

# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 110061

TO: Georgia L Helmer  
Location: 9d14 / 9e12  
Wednesday, December 17, 2003  
Art Unit: 1638  
Phone: 308-7023  
Serial Number: 09 / 805550

From: Jan Delaval  
Location: Biotech-Chem Library  
CM1-1E07  
Phone: 308-4498

jan.delaval@uspto.gov

### Search Notes

*OK  
pending removed  
12/17/03 glh*

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:16:40 ; Search time 18.3376 Seconds  
(without alignments)  
2123.954 Million cell updates/sec

Title: US-09-805-550-2

Perfect score: 2036

Sequence: 1 MCLNVKTLKGTNFEIASPD.....BELTANYLLDHGHEFDQQQ 405

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1187.5	58.3	382	2 T14336	RAD23 protein, iso
2	1127	55.4	392	2 T04150	RAD23 protein homo
3	925.5	45.5	379	2 T14337	RAD23 protein, iso
4	840.5	41.3	367	2 F96827	protein F20B17.8 [
5	632	31.0	409	2 S44346	RAD23 protein homo
6	623	30.6	385	2 JC7783	RAD 23B protein -
7	588	28.9	368	2 T40115	uv excision repair
8	573	28.1	363	2 S44443	RAD23 protein homo
9	543	26.7	246	2 G86296	T24D18.27 protein
10	483.5	23.7	398	2 S50507	excision repair pr
11	436.5	21.4	372	2 T27774	hypothetical prote
12	361.5	17.8	113	2 H86296	F309.1 protein - A
13	229.5	11.3	551	2 C84549	probable ubiquitin
14	211	10.4	536	2 B84549	probable ubiquitin
15	202.5	9.9	502	2 T18562	hypothetical prote
16	193	9.5	142	2 T51479	hypothetical prote
17	168	8.3	373	2 S54583	ubiquitin-like pro
18	154.5	7.6	354	2 T38404	yeast dsk2 homolog
19	150.5	7.6	507	2 T44768	antifreeze glycope
20	150.5	7.4	581	2 T22341	hypothetical prote
21	149.5	7.3	1040	2 T29092	TSC-22 protein hom
22	147.5	7.2	424	2 T33663	hypothetical prote
23	147	7.2	292	2 T18584	F15C11.2 - Caenorh
24	146	7.2	681	2 A83455	DNA polymerase sub
25	144.5	7.1	852	2 T46091	hypothetical prote
26	144.5	7.1	907	2 AD2931	cell division prot
27	144.5	7.1	910	2 H98331	cell division prot
28	144	7.1	638	1 XXAV	dihydroliipoamide S
29	141.5	6.9	833	2 AP2089	hypothetical prote

30 141.5 6.9 2082 2 T37056  
31 141 6.9 347 2 H75253  
32 141 6.9 590 1 T35297  
33 141 6.9 647 2 T39141  
34 140.5 6.9 1116 2 S77213  
35 140 6.9 638 2 H82690  
36 140 6.9 2440 2 S39162  
37 140 6.9 2441 2 S39161  
38 139.5 6.9 1039 2 A12884  
39 139.5 6.9 1132 2 A35098  
40 139 6.8 340 2 A35630  
41 138.5 6.8 762 2 H87302  
42 138 6.8 421 2 A60058  
43 137.5 6.8 547 2 H83018  
44 137.5 6.8 865 2 A47282  
45 137.5 6.8 873 2 A47283

ALIGNMENTS

RESULT 1

T14336

RAD23 protein, isoform I - carrot

C:Species: Daucus carota (carrot)

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000

C:Accession: T14336

R:Sturm, A.; Leinhardt, S.

Plant J. 13, 815-821, 1998

A:Title: Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in yeast.

A:Reference number: Z17989; MUID:98345997; PMID:9681019

A:Accession: T14336

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-382 <STU>

A:Cross-references: EMBL:Y12013; NID:g1914682; PIDN:CAA72741.1; PID:g1914683

A:Experimental source: subspecies Queen Anne's Lace, isolate W001C

C:Genetics:

A:Gene: RAD23-1

C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 58.3%; Score 1187.5; DB 2; Length 382;  
Best Local Similarity 62.7%; Pred. No. 2.9e-64;  
Matches 255; Conservative 46; Mismatches 75; Indels 31; Gaps 11;

Qy 1 MCLNVKTLKGTNFEIASPDASVADVKRIITETGQSTYRADQOMLIYQGIKDETTLE 60  
Db 1 MKIYVTKLGSQFEIQVNPDDSVADVKRSIETAGAAVYPAQQOMLIYQGIKDETTLL 60  
Qy 61 SNGVAENSFLVIMLSKAKASSSGASTATTA---KAPATLAQPAAPVAPASVARTPTQAP 117  
Db 61 ENNVAENSFLVIMLSKSKSPSGSTTSTAAPKAPOTSAPPSPV-APAVS-----QPP 113  
Qy 118 VATAET-AP-PSVQPOQAPAAATVAATDDADVYSAASNLVFGNNLEQITQIILDMGGGTW 175  
Db 114 ASTLPVPAPAPAPATAPISAAVGSSEANVDSAAALLVAGSNLEGAIQIILDMGGGTW 173  
Qy 176 ERTTVVRLAAAYNNPERAIDYLSGIPENVEQPVAPAPAGQQTNOQASPA-QPAPA 234  
Db 174 DRDTVIRIVRAAFNNPERAVEYLSGIPQEAQAPPVAPSPSG-----QAANPLDQPPAA 228  
Qy 235 LPQVPSPASAGPNANPLNLPQGVPSGSGNPGVVPAGSGALDALROLPOFOALLQVQA 294  
Db 229 --AQPAPASAGPNANPLDLFPQGLPDMGSN-----AAGAGNLDFLRNQQFOALRAVQVS 281  
Qy 295 NPQILQMLQELGKQNPQILRLIQENQAEFLRLVNSPEGGPGGNILGQAAAAVPTQLTV 354  
Db 282 NPQILQMLQELGKQNPQILRLIQEQHQAQDFLQINPEMGE--NLLGH-----GPOAISV 335  
Qy 355 TPBERAIOBLEGNGFNRELVVFFACNKGDELTANYLLDHGHEFD 401  
Db 336 TPBERDAIERLEANGFDRELVLVFFACNKGNEELANYLLDHMHEFE 382

pending  
submitted  
8/21/03



Query Match	45.5%;	Score 925.5;	DB 2;	Length 379;
Best Local Similarity	50.2%;	Pred. No. 1.6e-48;		
Matches	203;	Conservative	57;	Mismatches 117; Indels 27; Gaps 8;
Qy	1	MKLNVTKLTGTFEIEASPDASVADVKRIIETTTGGSTYRADDOOMLIYOGKILKDETTLE	60	
Db	1	MKLTVTKLTGSHFEIRAPNDVTYMAIKONIEDLQGGKNDYPCGQOLLIHNGKVLKDSSTLA	60	
Qy	61	SNQVAENSLFVIMLSKAKA - SSSGASTATTAKAPATLAQPAAPVAPAAASVARTPTQAPVA	119	
Db	61	ESKISEDGLFVLMGSKTMSSTGTTPAAQSSSAPATPPAPAVAPAPAPAPASAVIPNT	120	
Qy	120	TATAPPSPVOQAPAAATVAATDDADYVSOAANLVFGNNLEOTIQOILDMGGGTWERDT	179	
Db	121	TVPEAP--LSPAPAP-----SDTYGEAASNNVAGSLNEQTIOHIMDMGGGMDTNM	169	
Qy	180	VVRALRAAYNNPRAIDYLSGIPENVEAQ - PVARAPAAQGGTNNQQAASPAQPAVALPVQ	238	
Db	170	VSRALRAAYNNPRAVDYLSGIPENAAAVFVSH--FGDQINAGNNAISDNGVA----	223	
Qy	239	PSPASAGPNANPLNLPQGVPGSGSNPGVVPFGAGSGALDALRLOPQALLQLVQANPQI	298	
Db	224	-GAAPGAPNSLPLNMPFPQETLSG-----VTGAGLSLEFLRNPQFOTLRSVMVQRNPQI	276	
Qy	299	LQPMLOELGKONQOILRLTOENQAEFLRLVNESPEGPGVNIQLGOLAAAVPQTLTVTPEE	358	
Db	277	LQPMLELGLKONPOLLRQIQEHHEBFQLQINERVEASE-GDMFDQEPQDVPQBITVTAAAD	335	
Qy	359	REAIQRLGFMGNRELVLVFFACNKDDELTNYLLDHGHEPDD	402	
Db	336	QEAIERLEANGFDRGLVIEAFLACDRNEELAVNYLLENAGDFED	379	
RESULT 4				
F96827		protein F20B17.8 [imported] - Arabidopsis thaliana		
C:Species:		Arabidopsis thaliana (mouse-ear cress)		
C:Date:		02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002		
C:Accession:		F96827		
R:Theologis, A.; Ecker, J. R.; Palm, C. J.; Federspiel, N. A.; Kaul, S.; White, O.				
Chin, C. W.; Chung, M. K.; Conn, L.; Conway, A. B.; Conway, A. R.; Creasy, T. H.; D.				
ansen, N. F.; Hughes, B.; Huizar, L.				
Nature		408, 816-820, 2000		
A:Authors:		Hunter, J. L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.		
C.A.; Li, J. H.; Li, Y.; Lin, X.; Liu, S. X.; Liu, Z. A.; Luros, J. S.; Maiti, R.;				
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.				
A:Authors:		Salzberg, S. L.; Schwartz, J. R.; Shinn, P.; Southwick, A. M.; Sun, H.;		
ker, M.; Wu, D.; Yu, G.; Fraser, C. M.; Venter, J. C.; Davis, R. W.				
A:Title:		Sequence and analysis of chromosome 1 of the plant Arabidopsis.		
A:Reference number:		A86141; MUID:21016719; PMID:11130712		
A:Accession:		F96827		
A>Status:		preliminary		
A:Molecule type:		DNA		
A:Residues:		1-367 <STO>		
A:Cross-references:		GB:AE005173; NID:g7715605; PIDN:AAF68123.1; GSPDB:GN00141		
C:Genetics:				
A:Gene:		F20B17.8		
A:Map position:		1		
C:Superfamily:		unassigned ubiquitin-related proteins; ubiquitin homology		
Query Match	41.3%;	Score 840.5;	DB 2;	Length 367;
Best Local Similarity	47.4%;	Pred. No. 1.9e-43;		
Matches	192;	Conservative	53;	Mismatches 119; Indels 41; Gaps 8;
Qy	1	MKLNVTKLTGTFEIEASPDASVADVKRIIETTTGGSTYRADDOOMLIYOGKILKDETTLE	60	
Db	1	MKLTVTKLTGSHFEIRVLPSTIMAVKKNIEDSGKNDYPCGQOLLIHNGKVLKDETSLV	60	
Qy	61	SNQVAENSLFVIMLSKAKASSGASTATTAKAPATLAQPAAPVAPAAASVARTPTQAPVA	120	
Db	61	ENKVTESGLFVIMLSKSKSGGS-----AQASQVQCVRLLLFHSLFPLPHELRSSINYPVFT	116	

QY 121 AETAPSVQQAAPAAATVADDDVYVSOAASNLVFGNNLEQTIQOILDMGGTWERDVT 180  
 Db 117 CSCFTYS-----CSTTGTDTYGAASTLVSGSSLEQVMVQOIMEGGSGWMDKTV 166  
 QY 181 VRALRAAYNNPERAIDYLYSGIPENVE-AQPVARAPAAGQQTNOQAASPAQPAVALPVQP 239  
 Db 167 TRALRAAYNNPERAIDYLYSGIPOTAEVAVPPEAQIAG-----SGAAPV-- 211  
 QY 240 SPASAGNANPLNLFQGVPSGSGNPGVVGAGSGALDALRQI-POFQALLQVLQVQANPOI 298  
 Db 212 APASGGPNSPLDLFFQ-----ETVAAGSGDLGTLFLRNNDQVAILTISAFSLN 262  
 QY 299 LQPMLELQKONPQIILRQENQAEFLRLVNESPEGGPG-GNILGQLAAAVPQTLTVTPE 357  
 Db 263 CEPMLQELKONPQIILRQENQAEFLRLVNEPEGSGDGEQDFQPEQEMPHAINVTFA 322  
 QY 358 ERAIORLEGMGNRLVLEVPFACNKBELTANYLLDHGHEPDD 402  
 Db 323 EQEAIORLEMGFDRLVIEAFIACDRNEELANYLLENSGDPED 367  
 RESULT 5  
 S44346  
 RAD23 protein homolog - human  
 N/Alternate names: p58/HHR23B; XP-C repair complementing protein  
 C/Species: Homo sapiens (man)  
 C/Date: 25-Dec-1994 #sequence\_revision 01-Sep-1995 #text\_change 24-Sep-1999  
 C/Accession: S44346  
 R/Masutani, C.; Sugawara, K.; Yanagisawa, J.; Sonoyama, T.; Ui, M.; Enomoto, T.; Takio, EMO J. 13, 1831-1843, 1994  
 A/Title: Purification and cloning of a nucleotide excision repair complex involving the A/Reference number: S44345; MUID:94222030; PMID:8168482  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-409 <WAS>  
 A/Cross-references: EMBL:D21090; NID:g498147; PIDN:BAA04652.1; PID:d1005181; PID:g498148  
 C/Superfamily: unassigned ubiquitin-related proteins; ubiquitin homolog  
 F/1-80/Domain: ubiquitin homolog <UBH>  
 Query Match 31.0%; Score 632; DB 2; Length 409;  
 Best Local Similarity 34.1%; Pred. No. 7.4e-31; Mismatches 138; Indels 70; Gaps 11;  
 Matches 150; Conservative 82; Mismatches 138; Indels 70; Gaps 11;  
 QY 1 MKNLVTKLTGNTFEIASPDASVADVVKRIETTQGSTYRADQOQLIYQKILKDETTLE 60  
 Db 1 MQVTLTKLQOQTFKIDIDPEETVKALKKEIESEKDAFPVACQKLIYAGKILNDDTALK 60  
 QY 61 SNGVAENSFLVIMLSKAKASSSGASTATTAKAPA-----TLAQAAPVAPAAV 109  
 Db 61 EYKIDKFNFVVMVTKPKAVSTPAPATTQCSAPASTTAVTSSTTTVAQAQTPV-PALAP 119  
 QY 110 ARTPTQ-----APVATAETAPSVQQAAPAAAT-VNATDD-----ADVYSQ 149  
 Db 120 TSPFASITPASATSEAPASAAKQEPKAEKPAETVATSTPATDSTSGDSRSRLFED 179  
 QY 150 AASNLVFGNNLEQTIQOILDMGGTWERDVTVRALRAAYNNPERAIDYLYSGIPENVEAQ 209  
 Db 180 ATSALTVGSGYENMVTIEMSG---YERQVVAALRASFPNDRAVEYLLTGPVSEGNVGASDPA 236  
 QY 210 PVARAPAAGQQTNOQAASPAQPAALVFPVSPASAGNANPLNLFQGVPSGSGNPGVVP 269  
 Db 237 AVVDPFQA-----STGAQSSAVAAAAATTATT-----TTTSSGHP----- 275  
 QY 270 GAGSGALDALRQLPQFALLQVLQVQANPQILQMLQELKONPQILRIQENQAEFLRLVN 329  
 Db 276 -----LEFLRNQPFQOMRQITIQONPSLLPALLQOIGRENPLLQOISSHQEQFIQMLNEPAQEGQGGG 323  
 QY 330 E-----SPGGPGGNILGQLAAVQ-----QTLTVTPEERAIQRLGCMGNRELVLVFFFA 381  
 Db 330 EPVQEAAGGGGGGGGGGGGIAEGSGHMNYIQVTPOEKEAERLKAUGFPPEGLVQIAYFA 389  
 QY 382 CNKDELTANYLLDHGHEFD 401

Db 390 CEKNENLAANFLLQQNFDED 409  
 RESULT 6  
 JC7783  
 RAD 23B protein - channel catfish  
 C/Species: Ictalurus punctatus (channel catfish)  
 C/Date: 02-Apr-2002 #sequence\_revision 02-Apr-2002 #text\_change 02-Apr-2002  
 C/Accession: JC7783  
 R/Liu, Z.; Li, P.; Kocabas, A.; Karsi, A.; Ju, Z. Biochem. Biophys. Res. Commun. 289, 317-324, 2001  
 A/Title: Microsatellite-containing genes from the channel catfish brain: Evidence of tr A/Reference number: JC7783  
 A/Contents: Brain  
 A/Accession: JC7783  
 A/Molecule type: mRNA  
 A/Residues: 1-385 <LIU>  
 C/Comment: This protein with a polythreonine tract, has importance in the nucleotide ex C/Genetics:  
 A/Gene: rad23b  
 A/Introns: 76/3  
 Query Match 30.6%; Score 623; DB 2; Length 385;  
 Best Local Similarity 34.6%; Pred. No. 2.4e-30; Mismatches 147; Conservative 85; Mismatches 129; Indels 64; Gaps 11;  
 QY 1 MKNLVTKLTGNTFEIASPDASVADVVKRIETTQGSTYRADQOQLIYQKILKDETTLE 60  
 Db 1 MQVTLTKLQOQTFKIDIDAEETVKALKKEIENKESFPVAGQKLIYAGKILNDDTALK 60  
 QY 61 SNGVAENSFLVIMLSKAKASSSGASTATT-AKAPATLAQPAAPVAP--AASVARTPTQAP 117  
 Db 61 EYKIDKFNFVVMVAKPKAATAAQAQSSSTTAASSSSSTSTTTTTPVPPVAASAATTTTTT 120  
 QY 118 VAT-----AETAPPVQQAAPAAATVADDDVYVSOAASNLVFGNNLEQT 163  
 Db 121 TTTTSTSESSVIEKAAEEKPPSTPASGSLT-----NVNIFEATSALTVGSYEMM 175  
 QY 164 IQOILDMGGTWERDVTVRALRAAYNNPERAIDYLYSGIPENVEAQPAAPAAQQTNQ 223  
 Db 176 VTEIMLMG---YERQVVAALRASFPNDRAVEYLLTGPVSEGNVGASDPA----- 226  
 QY 224 QAASPAQPAVALPVQPSPASAGNANPLNLFQGVPSGSGNPGVVPVFGAGSGA--LDALRQ 281  
 Db 227 -----PVGGAFA-----VTTGLSSPSTTPTQTAGSGANPLEFLRN 263  
 QY 282 LPQFALLQVLQVQANPQILQMLQELKONPQILRIQENQAEFLRLVN-----ESPEGGPG 337  
 Db 264 QPQFLQMRQITIQONPSLLPALLQOIGRENPLLQOISSHQEQFIQMLNEPAQEGQGGG 323  
 QY 338 GNILGQLAAAVPQTLTVTPEERAIQRLGCMGNRELVLVFFACNKBELTANYLLDHG 397  
 Db 324 G--VSEAGGGMNYIQVTPOEKEAERLKAUGFPPEGLVQIAYFACEKNLAANFLLQ-- 379  
 QY 398 HEFDD 402  
 Db 380 QNFDD 384  
 RESULT 7  
 T40115  
 uv excision repair protein rad23 homolog - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 02-Sep-2000  
 C/Accession: T40115; T51298  
 R/Wood, V.; Rajandream, M.A.; Barrell, B.G.; Taylor, K.; Harris, D. submitted to the EMBL Data Library, September 1998  
 A/Reference number: Z21906  
 A/Accession: T40115  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-368 <WOO>

A;Cross-references: EMBL:AL0311788; PIDN:CAA21170.1; GSPDB:GN00067; SPDB:SPBC2D10.12  
A;Experimental source: strain 972h-; cosmid c2D10  
R;Zhao, Y.; Elder, R.T.  
submitted to the EMBL Data Library, July 1999  
A;Description: A fission yeast orthologue (rhp23) of the human nucleotide excision repair  
A;Reference number: Z25362

A;Accession: T51298  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-368 <ZHA>  
A;Cross-references: EMBL:AF174293; PIDN:AAD51975.1  
A;Experimental source: strain SP223  
C;Genetics:

A;Map position: 2  
A;Gene: rhp23; SPDB:SPBC2D10.12  
A;Introns: 23/3; 48/1; 328/3; 351/3  
C;Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 28.9%; Score 588; DB 2; Length 368;  
Best Local Similarity 35.3%; Pred. No. 2.8e-28;  
Matches 146; Conservative 72; Mismatches 138; Indels 58; Gaps 13;

Qy 1 MKNVTKLGNFPI-BASPDASVADVVKRIIETTGQSTVRADOQMLIYQGIKDKETTL 59

Db 1 MNLTFKNLQOQKFVISDSVADTKISELKEKIQT---QQNYEVERQKLIYSGRIILADDKTV 57

Qy 60 ESNVGAENSLVIMLSKAKASSGASTATTAKAPATLAQAPAPVAPASVARTPTQAPVA 119

Db 58 GEYNKEQDFIVCWSRPKTS-----TPKSAASPAP--NPPASVPEKKVEAPSS 106

Qy 120 TAETAPSVQQAAPAT---VAATDDADVYSQAASNLVFGNNLEQTIQIILDMGGGTWE 176

Db 107 TVAESTSTQIVAAAPSPNPTTATSEAPI---DANTLAVCAQNVAVENNVENG---YE 160

Qy 177 RDTVVRAALRAAYNNPERAIDYLYSGIPENV---BAQPVAPAPAAQQTNOQAASPAQV 233

Db 161 RSEVERAMRAAFNPNPDRAVEYLLTGIPEDILNRQRESAALAAQQQOSEALA----- 213

Qy 234 ALPVQPSASAGPNANPLNLPQGVPSGSGNPGVVG-AGSGALDALRQLPQFQALLQLV 292

Db 214 -----PTSTG---QPANLFEQAALSENQEPSNTVGDPLGLFURSIPOQLRQIV 263

Qy 293 QANPOILQPMQLQELGKONPQILRIQENQAEFLRLVNESPEGG---PGNITLGQLAAVP 349

Db 264 QONPOMLETILQIQGGDPALAAQAITQNPEAFLLAEGESALPSGGI----- 314

Qy 350 QTLVTTPPERAIQRLGEGNRELVLVFFACNKEDELTANYLLDHGHEFDDQ 403

Db 315 -QIQITQESSEIDRLCQLGFDNRNIVIQAYLACDKNEELAANYLFEHGHESEDE 367

## RESULT 8

S4443

RAD23 protein homolog2 - human

N;Alternate names: HHR23A protein

C;Species: Homo sapiens (man)

C;Date: 25-Dec-1994 #sequence\_revision 01-Sep-1995 #text\_change 18-Feb-2000

R;Masutani, C.; Sugawara, K.; Yanagisawa, J.; Sonoyama, T.; Ui, M.; Enomoto, T.; Takio,

EMBO J. 13, 1831-1843, 1994

A;Title: Purification and cloning of a nucleotide excision repair complex involving the  
A;Reference number: S44345; MUID:94222030; PMID:8168482

A;Accession: S4443

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-363 &lt;MAS&gt;

A;Cross-references: EMBL:D21235; NID:g498145; PIDN:BA04767.1; PID:d1005299; PID:g498146

R;Lamerdin, J.; McCreedy, P.; Skilwagen, S.; Ramirez, M.; Carrano, A.

submitted to the EMBL Data Library, November 1996

A;Description: Characterization by genomic sequence analysis of a gene-rich 111 kb region

A;Reference number: Z22906

A;Accession: T45076

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-363 <LAM>

A;Cross-references: EMBL:AD000092; PIDN:AAB51177.1

A;Experimental source: cell line 5HL2-B; fibroblast

C;Genetics:

A;Gene: RAD23A

A;Map position: 19p13.2

A;Introns: 24/3; 78/3; 139/2; 158/1; 200/3; 227/1; 271/3; 326/3

C;Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

F;3-82/Domain: ubiquitin homology <UBH>

Query Match 28.1%; Score 573; DB 2; Length 363;

Best Local Similarity 33.3%; Pred. No. 2.2e-27;

Matches 139; Conservative 79; Mismatches 124; Indels 76; Gaps 13;

Qy 3 LNVTKLKGTFEASPDASVADVVKRIIETTGQSTVRADOQMLIYQGIKDKETTL 62

Db 5 ITLTKTQQOTFKIRMEPDETVKVLKKBKIBAKGRDAPPVAGQKLIYAGKILSDVPIRDY 64

Qy 63 GVAENSLVIMLSKAKASSGASTATTAKAPATLAQAPAA---PVAPASVARTPTQAPVA 119

Db 65 RIDEKFNVMVTKTKA---GQTSAPPEASPTAAPESSTSPFPATISGSHPP---PAA 118

Qy 120 TAETAP-----PSVQQAAPATVAATDDADVYSQAASNLVFGNNLEQTIQIILDMGGGT 174

Db 119 REDKSPSEASPTTSPESV-SGSVPSSGSGREDAASTLVTSGETMLETIMSMG--- 174

Qy 175 WERTVVRALRAAYNNPERAIDYLYSGIPENVEAQPVAPAPAAQO-QTNQQAASPAQV 233

Db 175 YERERVVAALRASAYNNPHRAVEYLLTGIPGSPE-----PEHGSVQESVSQAPATEA- 226

Qy 234 ALPVQPSASAGPNANPLNLPQGVPSGSGNPGVPGAGSGALDALRQLPQFQALLQLV 293

Db 227 -----AGNPLEFLRDLQFQFQNMQRQIV 249

Qy 294 ANPOILQPMQLQELGKONPQILRIQENQAEFLRLVNESPEGGPGNI-----LGQLAAA 347

Db 250 QNPALLPALLQQLQGOENPQLLOQISRHOEQFIQMLNEPP--GELADISDVEGEVGAIGEE 307

Qy 348 VPQT--LTVTPPERAIQRLGEGNRELVLVFFACNKEDELTANYLLDHGHEFDDQ 403

Db 308 APQWNYIQVTPQBEKAIERLKAIGFPELVQIYAFACERKNENLAANFL--SQNFDDDE 363

## RESULT 9

G86296

T24D18.27 protein - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001

C;Accession: G86296

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial;

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: G86296

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-246 &lt;STO&gt;

A;Cross-references: GB:AE005172; NID:g6587822; PIDN:AAF1851.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

Query Match 26.7%; Score 543; DB 2; Length 246;

Best Local Similarity 41.8%; Pred. No. 8.8e-26;

Matches 117; Conservative 41; Mismatches 74; Indels 48; Gaps 4;

**Query** 1 M K L N V K T L G N T F E I A S P D A S V A D V K R I I E T T Q G S T Y R A D Q O M L Y I Q G K I L K D E T T L 60  
**Db** 1 M K L T V K T L G S H F E I R V L P T D T I M A V K N G I E D S O K O N Y P C G Q Q L L I H N G K V L K D E T T V 60  
**Qy** 61 S N G V A E N S F I V I M L S K A G S S G A S T A T T A K A P A T T A Q P A A P A V A P A A S V A R T P T Q A P V A T 120  
**Db** 61 E N K V T E S G F L V W M L S K S T A S S A G P S T Q L A A P T T Q S I A V P A S N S T P V Q E O F T A - - - - - 115  
**Qy** 121 A E T A P P S V Q P Q A P A A N T V A A T D D A D V Y S Q A S N L I V F G N N L E Q T I Q I I L D M G G G T W E R D T V 180  
**Db** 116 - - - - - Q S D T Y G A A S T L V S G S S I E Q M V Q Q I M E M G G S W D K E T V 153  
**Qy** 181 V R A L R A A Y N N P E R A I D Y L S G I P E N V E A Q P V A R A P A G Q O T N Q O A A S P A Q P A V A L P V Q P S 240  
**Db** 154 T R A L R A A Y N N P E R A V D Y L Y G I P E T V - - - T I P A T N L S G V G S G R E L T A P - - - - - 198  
**Qy** 241 P A S A G P N A N P L N F P Q G V P S G S N P G V V P G A G S G A L D A L R 280  
**Db** 199 P P S G G P N S P L D L F P Q E A V S D A A - - - - - G G D L G T L E F L R 232

**RESULT 10**

S50507  
N:Alternate names: protein RAD23 - yeast (*Saccharomyces cerevisiae*)  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-Jan-1995 #sequence revision 12-May-1995 #text\_change 24-Sep-1999  
C:Accession: S50507; S30845; A54599; S38547  
R:Dietrich, F.S.  
submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of *S. cerevisiae* cosmids 8199, 8334, and 9871.  
A:Reference number: S50491  
A:Accession: S50507  
A:Molecule type: DNA  
A:Residues: 1-398 <DIE>  
A:Cross-references: EMBL:U18779; NID:g603625; PIDN:AAB65005.1; PID:g603642; MIPS:YEL037c  
R:Mulligan, J.T.; Dietrich, F.S.; Hennessy, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor, R.; Watkins, J.F.; Sung, P.; Prakash, L.; Prakash, S.  
submitted to the EMBL Data Library, February 1993  
A:Reference number: S30812  
A:Accession: S30845  
A:Molecule type: DNA  
A:Residues: 1-398 <MUL>  
A:Cross-references: GB:U18779; EMBL:L10830; NID:g603625; PIDN:AAB65005.1; PID:g603642  
R:Watkins, J.F.; Sung, P.; Prakash, L.; Prakash, S.  
Mol. Cell. Biol. 13, 7757-7765, 1993  
A>Title: The Saccharomyces cerevisiae DNA repair gene RAD23 encodes a nuclear protein  
A:Reference number: A54599; MUID:94067136; PMID:8246991  
A:Accession: A54599  
A:Molecule type: DNA  
A:Residues: 1-398 <WAT>  
A:Cross-references: GB:I25428; NID:g409246; PIDN:AAA16070.1; PID:g409247  
R:Meinick, L.; Sherman, F.  
J. Mol. Biol. 233, 372-388, 1993  
A>Title: The gene clusters ARC and COR on chromosomes 5 and 10, respectively, of *Saccharomyces cerevisiae*  
A:Reference number: S38543; MUID:94016558; PMID:8411151  
A:Accession: S38547  
A:Molecule type: DNA  
A:Residues: 1-276, 'R', 278-398 <MEL>  
A:Cross-references: EMBL:S66117; NID:g430823; PIDN:AAB28441.1; PID:g430824  
C:Genetics:  
A:Gene: SGD:RAD23  
A:Cross-references: SGD:S0000763; MIPS:YEL037c  
A:Map position: 5L  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology  
C:Keywords: DNA repair  
F:2-77/Domain: ubiquitin homology <UBH1>

**Query Match** 23.7%; Score 483.5; DB 2; Length 398;  
**Best Local Similarity** 32.3%; Pred. No. 5.9e-22;  
**Matches** 142; Conservative 65; Mismatches 144; Indels 89; Gaps 18;

**Qy** 1 M K L N V K T L G N T F E I A S P D A S V A D V K R I I E T T Q G S T Y R A D Q O M L Y I Q G K I L K D E T T L 59  
**Db** 1 M K L T V K T L G S H F E I R V L P T D T I M A V K N G I E D S O K O N Y P C G Q Q L L I H N G K V L K D E T T V 59

```

Db      2   VSLTFKNFKKKVPLDLLEPSNTILETK-----TKLAQSISCKESQIKLITYSGKVLDQSKTV 57
Qy     60   ESVGVAENSFLVIIMLSKAKASSGASTATTAKAPATLAQPAAPVAPAASVARTP-----113
          :|::||::||:||||:|::|::||:
Db    58   SECGLKGDDGVFMVSOKK-----STKTkvTEP-----PIAPES--ATTGRENST 101
          :|::||::||:||||:|::|::||:
Qy   114   -----TQPAVATASTAPPVOQAAPAAATVAATDADVVYSAASNLVFGNNLEQTIOQL 168
          :|::||::||:||||:|::|::||:
Db   102   EASPSTDAASAAPATAPEGSQFQEQTATTERTESA-----STPGFVVGTNETIERIM 156
          :|::||::||:||||:|::|::||:
Qy   169   DMGGTGWERDVVRALRAAYNYPERRAIDLYLSGIPENVEAQPVARAAPAGQQTNQOAAASP 228
          :|::||::||:||||:|::|::||:
Db   157   EMG---YQREEVERALRAAFNPNPRAVEYLMLGPENL-----RQPEPOQT---AAA 204
          :|::||::||:||||:|::|::||:
Qy   229   AQPAVALVPQSPSASAGNPANLNLPFGVPVGSSNGPVGPGAGSGALDALRLP-----283
          :|::||::||:||||:|::|::||:
Db   205   EQPSTAATTAEQPAED-----DLFAQAACQGNAASSCAL-GTTCCGATDAACGGPPGSG 256
          :|::||::||:||||:|::|::||:
Qy   284   ----QFOALLQIVQNPOILOPMQLQELGKQNPOILRLIQEINQAEFLRLVNES-----331
          :|::||::||:||||:|::|::||:
Db   257   LTVEDLLSLRWVSGNPEPALPLENISARYPOLREHITMANPEVFVSMLEAVGNMQDV 316
          :|::||::||:||||:|::|::||:
Qy   332   PEG-----GPGGNILGQAAAA-VPO-----TLTVTPEEREAIQRLEGMGFNRELVL 377
          :|::||::||:||||:|::|::||:
Db   317   MEGADDWVEGDIEIETVGEAAAGLGCGEGGESFQVDYTPEDDQAISRLCEUGFERDLV 376
          :|::||::||:||||:|::|::||:
Qy   378   VFACNKDBDELTANYLL-DH 396
          :|::||::||:||||:|::|::||:
Db   377   VFACDKNEEAANAILFSDH 396
          :|::||::||:||||:|::|::||:

RESULT 11
T27774
hypothetical protein ZK20_3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C:Accession: T27774
R:Gajadaty, S.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20417
A:Accession: T27774
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-372 <wtl>
A:Cross-references: EMBL:Z69904; PIDN:CAA93780.1; GSPDB:GN00020; CESP:ZK20.3
A:Experimental source: clone ZK20
C:Genetics:
A:Gene: CESP:ZK20.3
A:Map position: 2
A:Introns: 49/2; 73/3; 101/3; 369/3
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match       21.4%; Score 436.5; DB 2; Length 372;
Best Local Similarity 28.9%; Pred.No. 3.6e-19;
Matches 117; Conservative 71; Mismatches 120; Indels 97; Gaps 13

Qy   1   MKLVVKTLKGNFTETEASPDVASADVDRRIETTQGSTYRADQOMLIYGKLKDITLE 60
Db   52   LSVTRILTQNFNELNEDQTIAEVKALVASEKDD-YAPELQKLIYNKGIILDSRVKVG 110
          :|::||::||:||||:|::|::||:
Qy   61   SNGVAENSFLVIIMLSKAKASSGASTATTAKAPATLAQPAAPVAPAASVARPTQAPVAT 120
          :|::||::||:||||:|::|::||:
Db   111   EVGFDDSSKFVVVLMISKRKVTENVAPSSTV----ATAAEP-VFAAAA-----PASNPAPA 158
          :|::||::||:||||:|::|::||:
Qy   121   AETAPPSVQQAAPAAATVAATDDADVVYSAASNLVFGNNLEQTIOQLDMGGGTHERDTV 180
          :|::||::||:||||:|::|::||:
Db   159   ADVAPEA----AAPAEAALTDEQ-----EENVLAITGMG---YDREQT 195
          :|::||::||:||||:|::|::||:
Qy   181   VRALRAAYNYPERRAIDLYLSGIPENVEAQPVARAAPAGQQTNQOAASPAQPAVALPVQPS 240
          :|::||::||:||||:|::|::||:
Db   196   IALURAAFWNPDRVAFELUNGDPD-----AADQEPD-----227
          :|::||::||:||||:|::|::||:
Qy   241   PASAGNPANLNLPFGVPVGSSNGPVGAGSGALDALRLQFPQALLQLVQANPQILQ 300

```

Db 228 ---LGPEQNTIDN-----VDEGNDLNLNANMPQALAEIRALIQNPENIA 269

Qy 301 PMLQELKQNPQILRLIQENQAEFLRLVNESPEGGNIGLQAAAAVPQTLT----- 353

Db 270 AVLQQLAANPRLVQTQNNQQAQMDLLNGAQA-----GAAAGNAPERTPRRHVIH 323

Qy 354 VTPEREAIQRLEGMGFN--RELVLVFFACNKDELTANYLLDH 396

Db 324 LSPERAAAIERKAIVNAPEAVVVEYFACDKNEEAINFISN 368

RESULT 12

H86296

F3O9.1 protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001

C:Accession: H86296

R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso-Chan, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: H86296

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-113 <STO>

A:Cross-references: GB:AE005172; NID:94966345; PIDN:AAD34676.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 17.8%; Score 361.5; DB 2; Length 113;

Best Local Similarity 61.1%; Pred. No. 2.7e-15;

Matches 69; Conservative 22; Mismatches 21; Indels 1; Gaps 1;

Qy 291 LVQANPQILQMLQELGKQNPQILRLIQENQAEFLRLVNESPEGGPGG-NILGQLAAVVP 349

Db 1 MVNSNPQILQMLQELGKQNPQILRLIQENQAEFLRLVNESPEGGVDVDFDQDQEMP 60

Qy 350 QTLVTPEREAIQRLEGMGFNRELVLVFFACNKDELTANYLLDHGHEFDD 402

Db 61 HSVNVTPEQESIERLEANGFDRAIVIEAFVLCORNEELAAANYLLEHSADFED 113

RESULT 13

C84549

Probable ubiquitin-like protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: C84549

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: C84549

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-551 <STO>

A:Cross-references: GB:AE002093; NID:94584343; PIDN:AAD25138.1; GSPDB:GN00139

C:Genetics:

A:Map position: 2

Query Match 11.3%; Score 229.5; DB 2; Length 551;

Best Local Similarity 23.7%; Pred. No. 1.6e-06;

Matches 132; Conservative 61; Mismatches 170; Indels 193; Gaps 26;

Qy	3	LNVTKLGTNFIEASPDASVADVKRIIETTOGQSTYRADQQMLIYQGKILKDettLESN	62
Dd	20	VNIRCSNGTKFSVKTSLDSTVESFKELVAQS--SDVPANQQRLLIYKGRILKDDQTLLSY	76
Qy	63	GVAENSLFVLMSAKAKASSSGASTATTAKAPATLAQPAAPAVAPAASVARTPTQAPVATAE	122
Dd	77	GLQADH--TIHMVRGSAPSS-----APPAPAAQSOTTAPSIVTRGVGSD	117
Qy	123	TA-----PPSQPOAAPAATAVAATDDADV	147
Dd	118	NSSLGGASPGESLFPGLGFNPPLGGGNAMSLFGAGLPDLVTTQQLAQNPWMIRM-MN	176
Qy	148	SQAASNLFVGNLE-----QTIOQLDMG---GGTWERTDVVRALRAAYNNPE---	192
Dd	177	TPAIQNLM--NNPBFMRSMIMNPNQMRELVDRNPBLGHVLDPSILRQTLEAAARNPELMR	234
Qy	193	---RAIDLYXGIP-----ENVEAQPVARAAGAQTNTQQAASPAQPAV---	233
Dd	235	EMMENTORAMGIESMPEGFNRLMYENVQBELMNATTMSGNAGNNTGSPFAALLGNQ	294
Qy	234	ALPVQPPAS-----AG---PNANPLNLPQGVPSG-----CSNPGCV--	267
Dd	295	GVTTCGDSDANSSTPNAGTGTIPNANPLP-NPWGATGGQTTPAGRTNVGGDARSPLGG	353
Qy	268	VPGAGS-----GA-----LDLARQLPQFOALLQLVOANPQI-----L	299
Dd	354	LGLGLSLGGLGGLMGADSPLGATPFDAQSLSQLQNPALISQMWVSFNPQYMNQMSL	413
Qy	300	QPMLOELGKNPQIIURLIQENQAEFLRLVNESPE-----	333
Dd	414	NPQLRMLDSNPOLREMNO--NPDFLRQFS-SPENMQMMTLQOQSLSONRNTASQDAGOT	470
Qy	334	---GGPCG---NILQALAA---AVPQTLTVTPEREAA--IQLEGMGF-NRELVLV	378
Dd	471	GAATGNNGGLDLLNMNFGSLGAGLUGTGNSQSNVPPERYATQQLQEMGYDRAENIRA	530
Qy	379	FFACNKDELTANYILL	394
Dd	531	LLATGNVNAVERLL	546

```

RESULT 14
B84549
Probable ubiquitin-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84549
R:Liu, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; PMID:20083487; PIDN:AAD25137.1; GSPDB:GN00139
A:Cross-references: GB:AE002093; NID:g4584342;
C:Genetics:
A:Gene: At2g17190
A:Map position: 2

Query Match      10.4%; Score 211; DB 2; Length 536;
Best Local Similarity 21.5%; Pred. No. 2e-05;
Matches 125; Conservative 56; Mismatches 142; Indels 258; Gaps 23;

Qy    3 LNVTKLGTNFIEASPDASVADVKRIIETTOGQSTYRADQQMLIYQGKILKDettLESN 62
Dd   20 VNIRCSNGTKFSVKTSLDSTVESFKELIAQ---NSDVPANQQRLLIYKGRILKDDQTLLSY 76

Qy   63 GVAENSLFVLMSAKAKASSSGASTATTAKAPATLAQPAAPAVAPAASVARTPTQAPVATAE 122
Dd   77 GLQADH--TIHMVRGSAPSS-----APPAPAAQSOTTAPSIVTRGVGSD 117

Qy  123 TA-----PPSQPOAAPAATAVAATDDADV 147
Dd  118 NSSLGGASPGESLFPGLGFNPPLGGGNAMSLFGAGLPDLVTTQQLAQNPWMIRM-MN 176

Qy  148 SQAASNLFVGNLE-----QTIOQLDMG---GGTWERTDVVRALRAAYNNPE--- 192
Dd  177 TPAIQNLM--NNPBFMRSMIMNPNQMRELVDRNPBLGHVLDPSILRQTLEAAARNPELMR 234

Qy  193 ---RAIDLYXGIP-----ENVEAQPVARAAGAQTNTQQAASPAQPAV--- 233
Dd  235 EMMENTORAMGIESMPEGFNRLMYENVQBELMNATTMSGNAGNNTGSPFAALLGNQ 294

Qy  234 ALPVQPPAS-----AG---PNANPLNLPQGVPSG-----CSNPGCV-- 267
Dd  295 GVTTCGDSDANSSTPNAGTGTIPNANPLP-NPWGATGGQTTPAGRTNVGGDARSPLGG 353

Qy  268 VPGAGS-----GA-----LDLARQLPQFOALLQLVOANPQI-----L 299
Dd  354 LGLGLSLGGLGGLMGADSPLGATPFDAQSLSQLQNPALISQMWVSFNPQYMNQMSL 413

Qy  300 QPMLOELGKNPQIIURLIQENQAEFLRLVNESPE----- 333
Dd  414 NPQLRMLDSNPOLREMNO--NPDFLRQFS-SPENMQMMTLQOQSLSONRNTASQDAGOT 470

Qy  334 ---GGPCG---NILQALAA---AVPQTLTVTPEREAA--IQLEGMGF-NRELVLV 378
Dd  471 GAATGNNGGLDLLNMNFGSLGAGLUGTGNSQSNVPPERYATQQLQEMGYDRAENIRA 530

Qy  379 FFACNKDELTANYILL 394
Dd  531 LLATGNVNAVERLL 546

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:16:39 ; Search time 11.5265 Seconds  
(without alignments)  
1652.347 Million cell updates/sec

Title: US-09-805-550-2

Perfect score: 2036

Sequence: 1 MKLVNKLKGNFEIASPD.....EELTANYLLDHGHEFDQQQ 405

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	632	31.0	409	1 R23B_HUMAN	P54727 homo sapien
2	623.5	30.6	416	1 R23B_MOUSE	P54728 mus musculus
3	588	28.9	368	1 RH23_SCHPO	O74803 schizosacch
4	577	28.3	363	1 R23A_MOUSE	P54726 mus musculus
5	573	28.1	363	1 R23A_HUMAN	P54725 homo sapien
6	483.5	23.7	398	1 RA23_YEAST	P32628 saccharomyc
7	167	8.2	373	1 DSX2_YEAST	P48510 saccharomyc
8	154.5	7.6	354	1 YAUG_SCHPO	Q10169 schizosacch
9	149	7.3	1083	1 T2D3_HUMAN	O00268 homo sapien
10	144	7.1	637	1 ODP2_AZOV1	P10802 azotobacter
11	140.5	6.9	1509	1 GSRI_HUMAN	Q9NZM4 homo sapien
12	140	6.9	2441	1 CBP_MOUSE	P45481 mus musculus
13	139.5	6.9	1132	1 BAT3_HUMAN	P46379 homo sapien
14	137.5	6.8	547	1 ODP2_PSEAE	Q59638 pseudomonas
15	137.5	6.8	552	1 EFG1_CANAL	P43064 candida alb
16	137.5	6.8	865	1 CPN_DROME	Q02910 drosophila
17	137.5	6.8	1012	1 PHC1_MOUSE	Q64028 mus musculus
18	136	6.7	743	1 P021_HUMAN	P14859 homo sapien
19	136	6.7	797	1 PAT1_YEAST	P25644 saccharomyc
20	135	6.6	352	1 ALGP_PSEAE	P15276 pseudomonas
21	135	6.6	1048	1 SRA4_RAT	Q63627 rattus norv
22	134.5	6.6	699	1 VGLG_HSV2H	P13290 herpes simp
23	134	6.6	361	1 IF35_MOUSE	Q9DCB4 mus musculus
24	133.5	6.6	1157	1 SRA4_HUMAN	Q95104 homo sapien
25	133.5	6.6	2442	1 CBP_HUMAN	Q27993 homo sapien
26	131	6.4	1783	1 RAA3_CHLRE	Q9FEC4 chlamydomon
27	130	6.4	392	1 HME1_HUMAN	Q05925 homo sapien
28	129.5	6.4	1772	1 MSP1_PLAYO	P13828 plasmodium
29	129	6.3	511	1 P60_FISGR	Q01835 listeria gr
30	129	6.3	688	1 DP3X_HAEIN	P43746 haemophilus
31	129	6.3	1794	1 YRM5_SCHPO	Q10172 schizosacch
32	127.5	6.3	815	1 RCM5_HUMAN	P52756 homo sapien
33	127	6.2	497	1 HMES_DROME	P18488 drosophila

#### ALIGNMENTS

##### RESULT 1

R23B\_HUMAN  
ID R23B\_HUMAN STANDARD; PRT; 409 AA.  
AC P54727; Q8WUB0;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE UV excision repair protein RAD23 homolog B (HHR23B) (XP-C repair  
DB complementing complex 58 kDa protein) (P58).  
GN RAD23B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=94222030; PubMed=8168482;  
RA Masutani C., Sugawara K., Yanagisawa J., Sonoyama T., Ui M.,  
RA Konomoto T., Takio K., Tanaka K., van der Spek P.J., Bootsma D.,  
RA Hoeijmakers J.H.J., Hanaoka F.;  
RT "Purification and cloning of a nucleotide excision repair complex  
RT involving the Xeroderma pigmentosum group C protein and a human  
RT homologue of Yeast RAD23";  
RL EMBL J. 13:1831-1843 (1994).  
[2]  
RP SEQUENCE FROM N.A., AND VARIANT VAL-249.  
RN Rieder M.J., Livingston R.J., Daniels M.R., Montoya M.A., Chung M.-W.,  
RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,  
RA Schackwitz W.S., Sherwood J.K., Wirak L.A., Nickerson D.A.;  
RN Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RA Ramsay H.;  
RN Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
[4]  
RP SEQUENCE FROM N.A., AND VARIANT VAL-249.  
RN TISSUE=Uterus;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.M., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length

P31503 rattus norv  
Q9nrj4 homo sapien  
Q10690 mycobacteri  
P28284 herpes simp  
P13595 mus musculu  
Q96rn5 homo sapien  
Q24523 drosophila  
P35658 homo sapien  
P15205 rattus norv  
P03186 epstein-bar  
Q09472 homo sapien  
P21519 drosophila







SEQUENCE FROM N.A.

RC STRAIN=972;  
RC MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Scouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Borzym K., Langer I., Beck A., Lehner H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaie V., Motier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Rhode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,  
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=20119127; PubMed=10652237;  
RT Lombaerts M., Goeloe J.I., den Dulk H., Brandsma J.A., Brouwer J.;  
RA "Identification and characterization of the rhp23(+) DNA repair gene  
in Schizosaccharomyces pombe.";  
RL Biochem. Biophys. Res. Commun. 269:210-215(2000).  
CC -1- FUNCTION: Involved in postreplication repair of UV-damaged DNA.  
CC Postreplication repair functions in gap-filling of a daughter  
CC strand on replication of damaged DNA.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: Contains 1 ubiquitin-like domain.  
CC -1- SIMILARITY: Contains 2 UBA domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcement>  
CC or send an email to [license@sib.ac.uk](mailto:license@sib.ac.uk)).  
CC  
CC EMBL; AF174293; RAD51975.1; --  
CC EMBL; AL031788; CNA21170.1; --  
CC PIR; T40115; T40115.  
CC HSP; P54725; IDV0.  
CC GeneDB Spombe; SPBC2D10.12; --  
CC InterPro; IPR004806; Rad23.  
CC InterPro; IPR006636; STI1.  
CC InterPro; IPR000449; UBA\_domain.  
CC InterPro; IPR000626; Ubiquitin.  
CC Pfam; PF00627; UBA; 2.  
CC SMART; SM00727; STI1; 1.  
CC SMART; SM00165; UBA; 2.  
CC SMART; SM00213; UBA; 1.  
CC TIGRFAMs; TIGR00601; rad23; 1.  
CC PROSITE; P850030; UBA; 2.  
CC PROSITE; P850053; UBIQUITIN 2; 1.  
KW DNA damage; DNA repair; Nuclear protein; Repeat.  
FT DOMAIN 1 77 UBIQUITIN-LIKE.  
FT DOMAIN 135 185 UBA 1.  
FT DOMAIN 320 360 UBA 2.  
FT DOMAIN 119 122 POLY-ALA.  
FT

DR PROSITE; P850030; UBA; 2.  
DR PROSITE; P850030; UBIQUITIN 2; 1.  
KW DNA damage; DNA repair; Nuclear protein; Repeat.  
FT DOMAIN 1 79 UBIQUITIN-LIKE.  
FT DOMAIN 188 228 UBA 1.  
FT DOMAIN 371 411 UBA 2.  
FT DOMAIN 255 261 POLY-ALA.  
FT DOMAIN 262 270 POLY-THR.  
FT DOMAIN 336 355 POLY-GLY.  
SQ SEQUENCE 416 AA; 43516 MW; 1380245A6D892205 CRC64;  
Query Match 30.6%; Score 623.5; DB 1; Length 416;  
Best Local Similarity 32.4%; Pred. No. 2.5e-28;  
Matches 146; Conservative 85; Mismatches 136; Indels 83; Gaps 10;  
QY 1 M K L N V T L K G T N P E I E A S P D A S V A D V K R I I E T T G O S T V R A D Q O M L I Y Q G K I L K D E T T L E 60  
DB 1 M Q V T L K L Q O O F T K I D I D E E T V K A L K E I E K G K D A F P V A G O K L I Y A G K I L S D T A L K 60  
QY 61 S N G V A E N S F L V I M L S K A K A ----- S S G A S T A T T A K A P A T L A Q A P A P A A S V A 110  
DB 61 E Y K I D E K N F V V M V T K P A V T T A V P A T T Q P S S T P S F T T V S S P A V ----- A A A Q A P A P T P A 116  
QY 111 R T P Q T P A V A T ----- A E T A P ----- P S V O P Q A P A P A T V A T D ----- A D V 146  
DB 117 L A P T S P A S T T P A S T P A S E P A G A T O P K P A E K P A Q T P V L T S P A D A S T P G D S S R S N L 176  
QY 147 Y S O A S N L V F G N L E Q T I O Q I L D M G G T W E R D T V V R A L R A A V N N P E R A I D Y L S G I P E N V 206  
DB 177 F E D A T S A L V T G O S Y E N W V T E I M S G ----- Y E R Q V T A L R A S F N N P D R A V E L L M G I P G D R 233  
QY 207 E A Q V A R A P A A G Q T T N Q Q A S P A Q A P A V L P V O P S P A S A P N A N L P F P Q V P S G S N P G 266  
DB 234 E S Q A V D P P P Q A V S T ----- G T Q S P A V A A A A T T T A T T ----- T T T S G G H P - 275  
QY 267 V P G A S G S G A L D A L R Q L P Q A L L Q L V Q A N P Q I L Q P M L Q E L G K O N P Q I L R L T Q E N Q A F E L R 326  
DB 276 ----- L E F L R N Q P F Q M Q R I I Q N P S L P A L L Q I G R E N P Q L L Q Q I S Q B Q E H F I Q 326  
QY 327 L V N E S P E ----- G G P G N I L G L A A A V P Q T L V T P E E R A I Q R L E G M G F N 371  
DB 327 M L N E P V Q E A G G G G G G G G G G G G G G G G G G G A G S G H N Y I Q V T P Q E A L E R L K A L G F P 386  
QY 372 R E L I V F F A C K N D E L T A N Y L L D H G H E P D 401  
DB 387 E G L V I Q A Y F A C E K N E N L A A N F L L Q Q N F D E D 416  
RESULT 3  
RH23 SCHPO ID RH23 SCHPO STANDARD; PRT; 368 AA.  
AC 074803;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 15-SEP-2003 (Rel. 42, Last annotation update)  
DE UV excision repair protein rhp23 (RAD23 homolog).  
GN RHP23 OR SPBC2D10.12.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RN SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.  
RP STRAIN=SP223;  
RX MEDLINE=21648918; PubMed=11788722;  
RA Elder R.T., Song X.-Q., Chen M., Hopkins K.M., Lieberman H.B.,  
RA Zhao Y.;  
RT "Involvement of rhp23, a Schizosaccharomyces pombe homolog of the  
RT human HHR23A and Saccharomyces cerevisiae RAD23 nucleotide excision  
RT repair genes, in cell cycle control and protein ubiquitination.";  
RL Nucleic Acids Res. 30:581-591(2002).  
RN [2]

```
FT DOMAIN 205 208 POLY-GLN.
SQ SEQUENCE 368 AA; 4013 MW; 5CE75EB7E190EFD4 CRC64;

Query Match
Best Local Similarity 28.9%; Score 588; DB 1; Length 368;
Matches 146; Conservative 72; Mismatches 138; Indels 58; Gaps 13;

Qy 1 MKNLTKLGTNFEI-EASPDASVADVKRIETTTQGSTYRADQOQMLIYQKILKDETL 59
Db 1 MNLTKLQOQKFVISDVSDTKISELKEIQT---QOYEVERQKLIYSGRILADKTV 57

Qy 60 ESNQVAENSLVIMLSKAKASSGASTATTAKAPATLAQPAAPVAPASVARTPTQAPVA 119
Db 58 GEYNIKQDFIVCMVSPKSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 106

Qy 120 TAETAPSVQPAAPAT---VAATDDADVYQOASNLVFGNLEQTIQIILDMGGTWE 176
Db 107 TVAESTSTTQVAAAPSNPDPTTATSEAPI---DANTLAVGAQRNVAENVMENG---YE 160

Qy 177 RDTVTRALRAAYNNPERAIDLYSGIPENV---EAOPVARAPAAQOQTNQOASPAQPAV 233
Db 161 RSEVERAMRAAFNPDRAVEYLLTGIPEDILNQRSESAALAAQOQSEALA----- 213

Qy 234 ALPVQPSASAGNPANPLNFPQGVPGSGSNPGVPG-AGSGALDALROLPOFQALLQLV 292
Db 214 -----PTSTG---QANLFEQAALSENENQEPSNTVGDPLGLFLRSIPQFQQLRQIV 263

Qy 293 QANPQIILQMLQELKQNPQILRLIOENQAEFLRLVNESPEGG---PGNILQGLAAVP 349
Db 264 QONPQMLTLLQIQGGQDPAQAQAITQNPFAFLQLAAGAESALPSGGI----- 314

Qy 350 QTLVTTPPEERAIQRLGSGNRELVLVEFPACNKEELTANYLLDGHGFEFDDQ 403
Db 315 -QIQITQEESEDRLCQLGDFDRNIVIQAYLADCKNEELANYLFEHGHSEDEB 367

RESULT 4
R23A_MOUSE
ID R23A_MOUSE STANDARD; PRT; 363 AA.
AC P54726;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UV excision repair protein RAD23 homolog A (MHR23A).
GN RAD23A OR MHR23A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
STRAIN=BALB/c; TISSUE=Testis;
MEDLINE=86403997; Pubmed=8808275;
RA van der Spek P.J., Visser C.E., Hanaoka F., Smit B.,
RA Hagemeijer A., Bootsma D., Hoeijmakers J.H.J.;
RT "Cloning, comparative mapping, and RNA expression of the mouse
RT homologues of the Saccharomyces cerevisiae nucleotide excision repair
RT gene RAD23."
RL Genomics 31:20-27(1996).
CC -1- FUNCTION: Involved in postreplication repair of UV-damaged DNA.
CC Postreplication repair functions in gap-filling of a daughter
CC strand on replication of damaged DNA (Potential).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: Contains 1 ubiquitin-like domain.
CC -1- SIMILARITY: Contains 2 UBA domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; X92410; CAAG3145.1; --
DR HSSP; P54725; LDV0.
DR MGD; MGI:105126; Rad23a.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STI1.
DR InterPro; IPR000449; UBA domain.
DR Pfam; PF006626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00727; STI1; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS50030; UBA; 2.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
KW DNA damage; DNA repair; Nuclear protein; Repeat.
FT DOMAIN 1 79 UBIQUITIN-LIKE.
FT DOMAIN 161 201 UBA 1.
FT DOMAIN 318 358 UBA 2.
SQ SEQUENCE 363 AA; 39769 MW; 67EAB96EBE5A5C203 CRC64;

Query Match 28.3%; Score 577; DB 1; Length 363;
Best Local Similarity 33.4%; Pred. No. 8.6e-26;
Matches 140; Conservative 77; Mismatches 124; Indels 78; Gaps 13;

Qy 3 LNVKTLKGTNFEI-EASPDASVADVKRIETTTQGSTYRADQOQMLIYQKILKDETL 62
Db 5 ITLTKLQOQTFKIRMEPDETVKVLEKEIEAEKRDAPFVAGQKLIYAGKILSDVDVIRY 64

Qy 63 GVAENSLVIMLSKAKASSGASTATTAKAPATLAQPAAPVAP---AASVARTP----- 113
Db 65 HIDEKNFVVVMTKKA-CQGISAPPEA-SPTAVPEPSTFPFVLASGSHSPPTSRDK 122

Qy 114 TQAPVATAETAPSVQPAAPATVAATDDADVYQOASNLVFGNLEQTIQIILDMGG 173
Db 123 SPSEESTTTTSPESISGSPSSGSGREED-----AASLTVTGSEYETMLTEINSG-- 174

Qy 174 THERDTVPALRAAYNNPERAIDLYSGIPENVQAPVARAPAAQO-QTNQOASPAQPA 232
Db 175 -YERERKVAALRAASYNNPHRAVEYLLTGIPGSP-----PEHGSVQESRAEQPATEA 226

Qy 233 VALPVQPSASAGNPANPLNFPQGVPGSGSNPGVPGVAGSGALDALROLPOFQALLQLV 292
Db 227 -----AGENPLEFLRDPQFQNNRQVI 248

Qy 293 QANPQIILQMLQELKQNPQILRLIOENQAEFLRLVNESPEGGPGNI-----LGQLAA 346
Db 249 QONPALLPALQOQENPQLQISRHQEQFIQMLNEPP--GELADISDVEGEVGAIGE 306

Qy 347 AVPQT--LVTTPPEERAIQRLGSGNRELVLVEFPACNKEELTANYLLDGHGFEFDDQ 403
Db 307 EAPQMYIQTVPQKEAIERLKALGFPESLVIQAYFACEKNENLAANFULL--SQNFDD 363

RESULT 5
R23A_HUMAN
ID R23A_HUMAN STANDARD; PRT; 363 AA.
AC P54725;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UV excision repair protein RAD23 homolog A (HHR23A).
GN RAD23A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=94222030; Pubmed=8168482;
RA Masutani C., Sugawara K., Yanagisawa J., Sonoyama T., Ui M.,
RA Enomoto T., Takio K., Tanaka K., van der Spek P.J., Bootsma D.,
```

RA Hoesjmakers J.H.J., Hanaoka F.;  
RT "Purification and cloning of a nucleotide excision repair complex  
RT involving the Xeroderma pigmentosum group C protein and a human  
RT homologue of yeast RAD23.";  
RL EMBO J. 13:1831-1843(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Lamerdin J.E., McCready P.M., Stilwagen S., Ramirez M., Carrano A.;  
RT "Characterization by genomic sequence analysis of a gene-rich 111 kb  
RT region of 19p13.2 containing the human DNA repair gene, RAD23A.";  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP STRUCTURE BY NMR OF 319-363.  
RX MEDLINE=99061330; PubMed=9846873;  
RA Dieckmann T., Withers-Ward E.S., Jarosinski M.A., Liu C.F.,  
RA Chen I.S.Y., Feigon J.;  
RT "Structure of a human DNA repair protein UBA domain that interacts  
RT with HIV-1 Vpr.";  
RL Nat. Struct. Biol. 5:1042-1047(1998).  
RN [5]  
RP STRUCTURE BY NMR OF 319-363.  
RX MEDLINE=20541363; PubMed=11087358;  
RA Withers-Ward E.S., Mueller T.D., Chen I.S.Y., Feigon J.;  
RT "Biochemical and structural analysis of the interaction between the  
RT UBA(2) domain of the DNA repair protein HHR23A and HIV-1 Vpr.";  
RL Biochemistry 39:14103-14112(2000).  
CC -1- FUNCTION: Involved in postreplication repair of UV-damaged DNA.  
CC Postreplication repair functions in gap-filling of a daughter  
CC strand on replication of damaged DNA (Potential).  
CC -1- SUBUNIT: Interacts with MJD.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- DOMAIN: The ubiquitin-like domain mediates interaction with MJD.  
CC -1- SIMILARITY: Contains 1 ubiquitin-like domain.  
CC -1- SIMILARITY: Contains 2 UBA domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC -----  
DR EMBL; D21235; BAA04767.1; -;  
DR EMBL; AC000092; AAB1177.1; -;  
DR EMBL; BC014026; AAB14026.1; -;  
DR PIR; S44443; S44443.  
DR PDB; 1DVO; 11-FEB-00.  
DR PDB; 1P4I; 20-DEC-00.

DR PDB; 1IFY; 03-JUL-02.  
DR Genew; HGNC:9812; RAD23A.  
DR MIM; 600061; -;  
DR GO; GO:0003697; F: single-stranded DNA binding activity; TAS.  
DR GO; GO:0006289; P: nucleotide-excision repair; TAS.  
DR InterPro; IPR004806; Rad23.  
DR InterPro; IPR006636; STI1.  
DR InterPro; IPR000449; UBA domain.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00627; UBA; 2.  
DR Pfam; PF00240; ubiquitin; 1.  
DR SMART; SM00727; STI1; 1.  
DR SMART; SM00165; UBA; 2.  
DR SMART; SM00213; UBQ; 1.  
DR TIGRFAMs; TIGR00601; rad23; 1.  
DR PROSITE; PS00303; UBA; 2.  
DR PROSITE; PS00053; UBIQUITIN 2; 1.  
DR PROSITE; PS00053; UBIQUITIN 2; 1.  
DR DNA damage; DNA repair; Nuclear protein; 3D-structure; Repeat.  
FT DOMAIN 1 81 UBIQUITIN-LIKE.  
FT DOMAIN 161 201 UBA 1.  
FT DOMAIN 318 358 UBA 2.  
FT HELIX 321 324  
FT TURN 325 331  
FT HELIX 334 341  
FT TURN 342 345  
FT HELIX 348 354  
FT TURN 355 356  
SQ SEQUENCE 363 AA; 35609 MW; C4E47E9313BB47B5 CRC64;  
Query Match 28.1%; Score 573; DB 1; Length 363;  
Best Local Similarity 33.3%; Pred. No. 1.4e-25;  
Matches 139; Conservative 79; Mismatches 124; Indels 76; Gaps 13;  
QY 3 LNVKTLKGTNFEIEASPDASVADVKLIETQGSYTRADQOMLIYQKILKDETTLESN 62  
DB 5 ITLTKLQQQTFRMEPDETQVKVLEAEKGRDAPFVAGQKLIYAGKILSDDDVPIRDY 64  
QY 63 GVAENSLVIMLSKAKASSGASTATTAKAPATLAQPA--VPVPAASVARTTQAPVA 119  
DB 65 RIDEKPFVVVMVTKA---GGTSPAPPEASPTAEPSSSTFPAPTSGMHP--PAA 118  
QY 120 TASTP-----PSVQPAAPAAATVAATDDADVYSQAASNLVFGNNLEQTTIQIILDMGGGT 174  
DB 119 REDKSPSEASAPTPSPESV-SGSVPSSGSGREDAASTLVTSGEYETMLTSMG-- 174  
QY 175 WERTVTRALRAAYNNPERAIDLYLIGIPENVAQPARAPAQO-QTNQQAASPAQPAV 233  
DB 175 YERERVVAALRAAYNNPRAVEYLLTGIPGSPE-----PEHGSVQESQVSEQPATEA- 226  
QY 234 ALPVQSPASAGNPANLNLFPQVPSGSGNPVPGVPGSGGALDALRQLPQFQALLQVQ 293  
DB 227 -----AGENPLEFURDQFQFQNMQRQVIQ 249  
QY 294 ANPQILQPMQLQELGKONPQILRLIQENQAFRLVNESPEGGPGGNI-----LGQLAAA 347  
DB 250 QNPALLPALLQQLQGENPOLLQIISRHQEQIQLNLEPP--GELADISDVEGEVGAIEE 307  
QY 348 VPQT--LTVTPPEREAIQRLGEMGNRELVLVFFACNKKDELTANYLLDHGHEFDQO 403  
DB 308 APQNMVITVTPQEKEAIERLKALGFPELSLVIAQYFACEKNENLAANFL--SQNFDD 363  
RESULT 6  
RA23 YEAST STANDARD; PRT; 398 AA.  
ID RA23 YEAST  
AC P32628;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE UV excision repair protein RAD23.  
GN RAD23 OR YEL037C OR SYGP-ORF29.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;



CC -!- SIMILARITY: Contains 1 ubiquitin-like domain.  
CC -!- SIMILARITY: Contains 1 UBA domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L40587; AAB07267.1; -;  
DR EMBL: Z49704; CAA89774.1; -;  
DR PIR: S54583; S54583.  
DR HSSP: Q15843; 1NDD.  
DR SGD: S0004889; DSK2.  
DR InterPro: IPR006636; STI1.  
DR InterPro: IPR000449; UBA domain.  
DR InterPro: IPR000626; Ubiquitin.  
DR Pfam: PF00627; UBA; 1.  
DR Pfam: PF00240; ubiquitin; 1.  
DR PRINTS: PR00348; UBIQUITIN.  
DR SMART: SM00727; STI1; 2.  
DR SMART: SM00165; UBA; 1.  
DR SMART: SM00213; UBQ; 1.  
DR PROSITE: PS00303; UBA; 1.  
DR PROSITE: PS00299; UBIQUITIN\_1; 1.  
DR PROSITE: PS00553; UBIQUITIN\_2; 1.  
KW Nuclear protein.  
FT DOMAIN 1 76 UBIQUITIN-LIKE.  
FT DOMAIN 327 371 UBA.  
FT CONFLICT 109 109 R -> A (IN REF. 2).  
FT CONFLICT 296 296 R -> A (IN REF. 2).  
SQ SEQUENCE 373 AA; 39516 MW; 25EDF82B9DB67DF6 CRC64;  
  
Query Match 8.2%; Score 167; DB 1; Length 373;  
Best Local Similarity 23.2%; Pred. No. 0.0085;  
Matches 94; Conservative 47; Mismatches 168; Indels 96; Gaps 19;  
  
QY 1 MKNVTKLGTN-PEIEASPDASVADVKRIETTCGOSTYRADQQMLYQGIKLDKDTTL 59  
DB 1 MSLNIHKSGQDKWEVNVAPSTVLQFKAINKANG---IPVANORLIYSGKILKDDQTV 57  
  
QY 60 ESNQVAENFLVIMLSKAK---ASSGSASTATKAPATLAQAPAPVAPAAASVARTPTQA 116  
DB 58 ESYHIQDGHSHVLSKPKPQTASAAGANNATGCA-----AAGTGATPNWS 104  
  
QY 117 PVATAETAPSPVQPOAAPATVAATDADVYSQASNLVFGNNLEQTIOQILDMGGGTWE 176  
DB 105 S-QQSRGFNPLADLTSAARYAGVLYNMPADMFPGDGGAL---NNDSSNNQDELLRM---ME 156  
  
QY 177 RUTVVRLAAVNNPERADLYSGTPENVEAQPVAPAPAAQQTNOQAAP-----AQA 232  
DB 157 NPIFOSQNMELSNPO-MLDFMIQSNPQLQVAGPQARQLSQPMFQMLTNPDMIRQSNQ 215  
  
QY 233 VALPVQSPASAGPNANPLNLPFGVPSPG-----GSPNPGVVPAGGSA 275  
DB 216 FARWMDPN-AGWSAGGASAPF--APGDAPEEGSNTNTSSNTGNAGTNAGTNAGA 272  
  
QY 276 LDALRQLPQFALLQLVQANP---QILQPMQLBGLKQNPQILRLIQENQAEFLRLVNESPE 333  
DB 273 NTA-----ANPFASLLNPAL-----NP-----FANAGNRAST 299  
  
QY 334 GPGGNN--ILGQLAAVPTQL-----TVTPEER--EAIQLEGMGF 370  
DB 300 GMPAFDPALLASMFQPPVQASQAEDTRPPEERYEHQLRLQNDMGF 344  
  
RESULT 8  
YAUG\_SCHPO STANDARD; PRT; 354 AA.  
AC Q10169;  
DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Hypothetical protein C26A3.16 in chromosome I.  
GN SPAC26A3.16.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI TaxID=4896;  
RN [1]\_TaxID=4896;  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880 (2002).  
CC -!- SIMILARITY: TO YEAST DSK2  
CC -!- SIMILARITY: Contains 1 ubiquitin-like domain.  
CC -!- SIMILARITY: Contains 1 UBA domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z69240; CAA93239.1; -;  
DR PIR: T38404; T38404.  
DR GeneDB SPombe; SPAC26A3.16; -;  
DR InterPro: IPR006636; STI1.  
DR InterPro: IPR000449; UBA domain.  
DR InterPro: IPR000626; Ubiquitin.  
DR Pfam: PF00627; UBA; 1.  
DR Pfam: PF00240; ubiquitin; 1.  
DR SMART: SM00727; STI1; 2.  
DR SMART: SM00165; UBA; 1.  
DR SMART: SM00213; UBQ; 1.  
DR PROSITE: PS00303; UBA; 1.  
DR PROSITE: PS00553; UBIQUITIN\_2; 1.  
KW Hypothetical protein.  
FT DOMAIN 1 78 UBIQUITIN-LIKE.  
FT DOMAIN 309 353 UBA.  
SQ SEQUENCE 354 AA; 36819 MW; 1A99B2D97E73A831 CRC64;  
  
Query Match 7.6%; Score 154.5; DB 1; Length 354;  
Best Local Similarity 21.5%; Pred. No. 0.04;  
Matches 81; Conservative 58; Mismatches 178; Indels 59; Gaps 12;



```

QY 113 PTQAPVATETAPPVQQAAPAAATVAATDDADVTSQAASNLVFGNNLEQTIQIILDMGG 172
Db 268 PPPPPAPATLARPGEPAGPPTAAPAVPPPA-----AAQN-----G 304
QY 173 GTWERDVTVRALRAAYNNPERAIDLYSGIPENVEAQP-----VARAPAAGOOTNQ----- 223
Db 305 GS-----AGAAPAPAPA-----AGGPAAGVSGQPGPGGAAAPAPGVAESPVRKV 349
QY 224 QAASPAQPAVALPVQSPASA---GFNANPLMLFPQGV---PSGGSNPGVVPQAGSGALDA 278
Db 350 QAAPPAQTLAASGPASTAASWVIGPTWQGLPSPAAPVPPAPGPTGLPKGAAGAVTQS 409
QY 279 LRQLPQ-----FOALLQLVQANPQILQPMQLGKQNPQI---LRLIENQAEPFLRLVNE 330
Db 410 LSRTPATTGIRATLTPTLAPRLPQPQNPQNTQNFQLPQPMVLRSENGQLLMIPQ 469
QY 331 SPGGPGGNGILQALAA---AVPOTLTVTP 356
Db 470 A-----LAQMQAQAHAQPOT-TMAP 488

RESULT 10
ODP2_AZOV1
ID ODP2_AZOV1 STANDARD; PRT; 637 AA.
AC P10802;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase
DE complex (EC 2.3.1.12) (E2).
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 478;
RX MEDLINE=88271330; PubMed=3292237;
RA Hanemaaijer R., Janssen A., de Kok A., Veeger C.;
RT "The dihydrolipoyltransacetylase component of the pyruvate
RT dehydrogenase complex from Azotobacter vinelandii. Molecular cloning
RT and sequence analysis.";
RL Eur. J. Biochem. 174:593-599 (1988).
RN [2]
RP SEQUENCE OF 1-15 AND 380-415.
RX MEDLINE=88082750; PubMed=3691494;
RA Hanemaaijer R., de Kok A., Jolles J., Veeger C.;
RT "The domain structure of the dihydrolipoyl transacetylase component
RT of the pyruvate dehydrogenase complex from Azotobacter vinelandii.";
RL Eur. J. Biochem. 169:245-252 (1987).
RN [3]
RP LIPOYL DOMAIN CONFORMATION.
RX MEDLINE=89052887; PubMed=3191933;
RA Hanemaaijer R., Vervoort J., Westphal A.H., de Kok A., Veeger C.;
RT "Mobile sequences in the pyruvate dehydrogenase complex, the E2
RT component, the catalytic domain and the 2-oxoglutarate dehydrogenase
RT complex of Azotobacter vinelandii, as detected by 500 MHz 1H-NMR
RT spectroscopy.";
RL FEBS Lett. 240:205-210 (1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 381-637.
RX MEDLINE=92196586; PubMed=1549782;
RA Mattevi A., Obmolova G., Schulze E., Kalk K.H., Westphal A.H.,
RA de Kok A., Hol W.G.J.;
RT "Atomic structure of the cubic core of the pyruvate dehydrogenase
RT multienzyme complex.";
RL Science 255:1544-1550 (1992).
RN [5]
RP STRUCTURE BY NMR OF 1-78.
RX MEDLINE=94222112; PubMed=8068086;
RA Berg A., de Kok A., Vervoort J.;
RT "Sequential 1H and 15N nuclear magnetic resonance assignments and
RT secondary structure of the N-terminal lipoyl domain of the

```

```

RT dihydrolipoyl transacetylase component of the pyruvate dehydrogenase
RL complex from Azotobacter vinelandii.";
RN Eur. J. Biochem. 221:87-100 (1994).
RN [6]
RP STRUCTURE BY NMR OF 1-78.
RX MEDLINE=97234563; PubMed=9119000;
RA Berg A., Vervoort J., de Kok A.;
RT "Three-dimensional structure in solution of the N-terminal lipoyl
RT domain of the pyruvate dehydrogenase complex from Azotobacter
RL vinelandii.";
RN Eur. J. Biochem. 244:352-360 (1997).
CC -1- FUNCTION: The pyruvate dehydrogenase complex catalyzes the overall
CC conversion of pyruvate to acetyl-CoA and CO(2). It contains
CC multiple copies of three enzymatic components: pyruvate
CC dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and
CC lipoamide dehydrogenase (E3).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide = CoA + S-
CC acetyldihydrolipoamide.
CC -1- COFACTOR: THE E2 COMPONENT CONTAINS THREE COVALENTLY-BOUND LIPOYL
CC COFACTORS.
CC -1- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
CC SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
CC -1- SIMILARITY: Contains 3 lipoyl-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X12455; CAA30987.1; ALT_INIT.
DR PIR; S01017; XXAV.
DR PDB; IEAA; 31-OCT-93.
DR PDB; IEAB; 31-OCT-93.
DR PDB; IEAC; 31-OCT-93.
DR PDB; IEAD; 31-OCT-93.
DR PDB; IEAE; 31-OCT-93.
DR PDB; IEAF; 31-OCT-93.
DR PDB; ILYU; 12-MAR-97.
DR PDB; ILYV; 12-MAR-97.
DR PDB; IDPB; 20-APR-95.
DR PDB; IDPC; 20-APR-95.
DR PDB; IDPD; 20-APR-95.
DR InterPro; IPR001078; 2oxoacid_dh.
DR InterPro; IPR006256; Acef.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR004167; E3 binding.
DR InterPro; IPR003016; Lipoyl.
DR Pfam; PF00198; 2-oxoacid_dh; 1.
DR Pfam; PF00364; biotin_lipoyl; 3.
DR Pfam; PF02817; e3 binding; 1.
DR ProDom; PD001115; 2oxoacid dh; 1.
DR TIGRFAMs; TIGR01348; PDHac_trf_long; 1.
DR PROSITE; PS00189; LIPOYL; 3.
KW Glycolysis; Transferase; Acyltransferase; Repeat; Lipoyl;
KW 3D-structure.
FT INIT MET 0
FT DOMAIN 1 326 LIPOYL BINDING.
FT DOMAIN 327 380 E1/E3 BINDING.
FT DOMAIN 381 637 CATALYTIC.
FT BINDING 39 39 LIPOYL (POTENTIAL).
FT BINDING 156 156 LIPOYL (POTENTIAL).
FT BINDING 261 261 LIPOYL (POTENTIAL).
FT REPEAT 1 115 LIPOYL (POTENTIAL).
FT REPEAT 116 220
FT REPEAT 221 326
FT ACT_SITE 610 610 POTENTIAL.
FT STRAND 2 5
FT STRAND 12 18
FT TURN 22 23

```



FT STRAND 25 25  
 FT STRAND 31 36  
 FT STRAND 41 45  
 FT STRAND 51 55  
 FT TURN 59 60  
 FT TURN 62 64  
 FT TURN 65 66  
 FT STRAND 68 73  
 FT HELIX 403 406  
 FT HELIX 416 431  
 FT STRAND 434 442  
 FT STRAND 444 452  
 FT TURN 453 453  
 FT HELIX 454 459  
 FT TURN 460 461  
 FT HELIX 466 480  
 FT HELIX 482 484  
 FT STRAND 486 488  
 FT TURN 490 491  
 FT STRAND 495 497  
 FT STRAND 503 505  
 FT STRAND 507 509  
 FT TURN 510 511  
 FT STRAND 512 514  
 FT STRAND 517 518  
 FT HELIX 521 523  
 FT HELIX 526 541  
 FT TURN 542 543  
 FT HELIX 547 550  
 FT STRAND 555 559  
 FT TURN 561 563  
 FT TURN 574 575  
 FT STRAND 578 582  
 FT STRAND 586 591  
 FT STRAND 596 609  
 FT TURN 610 612  
 FT HELIX 615 630  
 FT HELIX 632 637  
 SQ SEQUENCE 637 AA; 64913 MW; D6063B4A5A385F84 CRC64;

Query Match 7.1%; Score 144; DB 1; Length 637;  
 Best Local Similarity 22.5%; Pred. No. 0.31;  
 Matches 80; Conservative 39; Mismatches 149; Indels 88; Gaps 11;

QY 19 PDASVADVRIET-TOGQSTYRADQMLYQKILKDETTLESNGVAENSLVIMLSKA 77  
 DB 122 PDIGSAGKARVIEVLVKAGDQVQAEQSLVLESKASMEIIPSPASGVVES--VAIQLNAE 179  
 QY 78 KASSSGASTATTAKAPATLAQPAAPVAPAAASVARTPTQAPVATAETAPPVQVQ----PQAA 133  
 DB 180 VGTGDLILTLRTTCA-----QAQPTAPAAAAAASPAPAPLAPAAAGPQEVKVPDICSAG 233  
 QY 134 PAATVAATDDADVYSQAASNLVFGNNLEQTIQIILDMGGGTWERDVTVVRALRAAYNNPER 193  
 DB 234 KARVIEVLVKAGDQVQAEQSLI-----VLESKASMEIIPSP 269  
 QY 194 AIDLYSGIENVEAQPVARAPAGAGQOTNOQAASPAQPAVALPVQSPASAGNANPLNL 253  
 DB 270 A-----AGVVESAVQVLAENVEGTGQDQILTLRVAGAAPSPGRARGSPGQAAAPGAAPAPA 324  
 QY 254 PQGVPSGGSPGVPVPGAGSGALDALRQLPQFQALLQVQANPQILQPMQLQELGKQNPQI 313  
 DB 325 -PVGAPSEN-----GAKVHAGPAVRQL-----AREFGVE----- 352  
 QY 314 LRLIQENQAEFLRLVNPSPGPGGNITLGQALAAVAPQTLVTTPEREAIORLEGMG 369  
 DB 353 -----LAAINST---GPRGRILKEDVQAVVKAMQKAKEAPAGAASGAG 394

RESULT 11  
 GSRI HUMAN  
 ID GSRI HUMAN  
 AC Q9NZM4;

STANDARD; PRT; 1509 AA.

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glioma tumor suppressor candidate region gene 1 protein.  
 GN GLTSCR1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
 RX MEDLINE=20175430; PubMed=10708517;  
 RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,  
 RA Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,  
 RA Scheithauer B.W., Louis D.N., Jenkins R.B.;  
 RT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor  
 region.";  
 RL Genomics 64:44-50(2000).  
 CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,  
 CC placenta, skeletal muscle, and pancreas, and at lower levels in  
 CC lung, liver, and kidney.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AF182077; AAF62874.1; -;  
 DR Genew; HGNC:4332; GLTSCR1.  
 DR MIM: 605690; -;  
 FT DOMAIN 37 45 POLY-GLY.  
 FT DOMAIN 884 889 POLY-PRO.  
 FT DOMAIN 1214 1225 POLY-SER.  
 FT DOMAIN 1282 1286 POLY-PRO.  
 FT DOMAIN 1294 1304 POLY-PRO.  
 SQ SEQUENCE 1509 AA; 152991 MW; 7C5144F443CE6821 CRC64;

Query Match 6.9%; Score 140.5; DB 1; Length 1509;  
 Best Local Similarity 23.5%; Pred. No. 1.3;  
 Matches 108; Conservative 36; Mismatches 135; Indels 181; Gaps 24;

QY 55 DETTLESNGVA-ENSLVIM--LSKAKASS-----SGASTATT-----AKAPA 94  
 DB 220 EPTLASAGVSPQAGLVQKNSAAVATLNGNSVFGGAGAAAPTGTGSPQPLAVAPG 279  
 QY 95 TLAQPAAPVAPAAASVARTPTQAPVATAETAPPVQV-----QPQAAPAAATVAATDDADVYSQ 149  
 DB 280 LGSSPLVP-APNVILHRTPTPIQPKPAGVLPKLYQLTKPKFAPAGATLTIQGEPCALPQ 338  
 QY 150 ----AASNLVFGNNLEQTIQIILDMGGGTWERDVTVR-----ALR-----AAY 188  
 DB 339 QPKAPQNLTF-----MAAGKAGQNVLSGFPAPALQANVFQKPPATTTGAAP 385  
 QY 189 NNPERAI-----DLYLSG-----IPENVEAQP----- 210  
 DB 386 POPPGALSXPMVSHLLNQSSSIVIPAQHMLPQGNQFLLPAPAVQLPQOLSALPANVGQ 445  
 QY 211 --VAPAPAAAGQ-----TNQQAASP-----AQPAV-- 233  
 DB 446 ILAAAPHTGGQLIANPILTNQNLGAPLSGLPVLAPHSGAHSNHLISNAPIQVGPALFQ 505  
 QY 234 -----ALPVQSPASAGPNANP-----LNLFPQGV-----PSGSGNPGVPGAGSGAL 276  
 DB 506 MPVSLAAGSLPTQSQPAPAGPAATTVLQGVTLPPSAVAMLTPDGLVQDPATPAATGEAA 565  
 QY 277 DALRQLPQFQALLQVQANPQILQPMQLQELGKQNPQILRLIQENQAEFLRLV-----NESP 332  
 DB 566 PVLTVQF-----APQAPPVAVSTPL--PLGLQQPQQAQ---QPPOAPTFOAAAPPOATTP 613  
 QY 333 EGGPG-----GNILQALAAVAPQTLTV-----TPEER 359







DR InterPro: IPR006256; AceP  
 DR InterPro: IPR000089; Biotin\_lipoyl.  
 DR InterPro: IPR004167; E3\_Binding.  
 DR InterPro: IPR003016; Lipoyl.  
 DR Pfam: PF00198; 2-oxoacid dh; 1.  
 DR Pfam: PF00364; biotin lipoyl; 2.  
 DR Pfam: PF02817; e3 binding; 1.  
 DR ProDom: PD001115; 2-oxoacid dh; 1.  
 DR TIGRFAM: TIGR01348; PDHac\_trf\_long; 1.  
 DR PROSITE: PS00189; LIPOYL; 2.  
 KW Glycolysis; Transferrase; Acyltransferase; Repeat; Lipoyl;  
 KW Complete proteome.  
 FT BINDING 41 LIPOYL (BY SIMILARITY).  
 FT BINDING 159 LIPOYL (BY SIMILARITY).  
 FT ACT\_SITE 520 POTENTIAL.  
 FT CONFLICT 225 A -> V (IN REF. 1).  
 FT CONFLICT 295 301 GGAGATG -> AVPAQR (IN REF. 1).  
 FT CONFLICT 328 329 MO -> IE (IN REF. 1).  
 SQ SEQUENCE 547 AA; 56709 MW; 24EL5CC9A590CB4 CRC64;

Query Match 6.8%; Score 137.5; DB 1; Length 547;  
 Best Local Similarity 23.2%; Pred. No. 0.59;  
 Matches 87; Conservative 50; Mismatches 119; Indels 119; Gaps 19;

QY 41 ADQQLIYCKILKDETTLESNGVAENSEFLVIMLSKAKASS-----GASTA 87  
 Db 30 ADQSLTLLE-----SKASME---IPSPKAGVVVKSIAKVGDTLKEGDEILEVEGGEQP 82  
 QY 88 TTKAKPATLQAPAAPVAPASVARTPTQAPVATAETAPSPVQOPAPAPAAATVAATDADVY 147  
 Db 83 AEAKAEAPAPAEAPKAE-----PAP-----APSEKPAAPAAASVQDIKVPDIG 128  
 QY 148 SOAASNLVFGNLEQTIQIILDMGGTWERDVTVRAL---RAAYNNPERAIDLYLSGIPE 204  
 Db 129 SAGKANVI-----EVMYKAGDTVEADQSLITLESKASMEIPSPA-----SGVVE 173  
 QY 205 NV-----EAQFVARAPAGQQTNOQAASPAQPAVALPVQSPASAGPNAN 249  
 Db 174 SVSIRKVGDEVTGDLILKLKVEGAAPAAEE---QPAAPAPAQ-AAAAPAAEQKPAAPAA 226  
 QY 250 PLNLFPQGVPSGSGNPGVPGAGSGALDALRQLPQFQALLQVQANPQILQPMLOELGKQ 309  
 Db 227 -----APAKADTPAPV---GAPSRGAK-----VHAGPAV-RMLAREFGVE 263  
 QY 310 NPQIL-----RLIQENQAEFLRLVNESPEGGPGGNILGQLAAAPQTLTTPTE----- 357  
 Db 264 LSEVKASGPKGRILKEDVQVFYKEQLQRAKSGAGATGG---AGIPP-----IPEVDFSKF 316  
 QY 358 ---EREAIQRLEGMG 369  
 Db 317 GEVEEVAMTRLMQVG 331

## RESULT 15

EFG1\_CANAL  
 ID EFG1\_CANAL STANDARD; PRT; 552 AA.  
 AC P43064;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Enhanced filamentous growth protein.  
 GN EFG1 OR EFG.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 10231;  
 RX MEDLINE=9729875; PubMed=9155024;  
 RA Stoldt V.R., Sonneborn A., Leuker C.E., Ernst J.F.;  
 RT "Efg1p, an essential regulator of morphogenesis of the human pathogen  
 Candida albicans, is a member of a conserved class of BHLH proteins

RT regulating morphogenetic processes in fungi.";  
 RL EMBL J. 16:1982-1991(1997).  
 CC -!- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR THAT STIMULATES  
 CC PSEUDOHYPHAL MORPHOGENESIS.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE EFG1/PHD1/STUA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z32687; CAA83640.1; ..  
 DR InterPro: IPR003163; Yeast\_DNA\_bind.  
 DR Pfam: PF02292; APSES; 1.  
 KW Transcription regulation; Nuclear protein; DNA-binding.  
 FT DOMAIN 34 37 POLY-GLN.  
 FT DOMAIN 46 50 POLY-GLN.  
 FT DOMAIN 83 93 POLY-GLN.  
 FT DOMAIN 315 326 POLY-ALA.  
 FT DOMAIN 332 338 POLY-PRO.  
 FT DOMAIN 438 446 POLY-GLN.  
 FT DOMAIN 470 473 POLY-GLN.  
 FT DOMAIN 483 493 POLY-GLN.  
 FT DOMAIN 496 499 POLY-GLN.  
 SQ SEQUENCE 552 AA; 59564 MW; F94FD94FC2E06EB7 CRC64;

Query Match 6.8%; Score 137.5; DB 1; Length 552;  
 Best Local Similarity 23.9%; Pred. No. 0.6;  
 Matches 80; Conservative 42; Mismatches 120; Indels 93; Gaps 18;

QY 79 ASSGA-----STATTAKAPATLQAPAPVAPASVAR--TPTQAPVATAETA-----P 125  
 Db 139 SAPSGAIPSNSTSGSPQPLPGQQAIVPIPHVSTWQQTPTVQDTLNASTSTVGQFQ 198  
 QY 126 PSVQPOQAAPA-----ATVAATDDAD---VYSQAASNLVFGNNL-----E 161  
 Db 199 PGIRPVTTTWDEKTLCYQVDANNVSVVRADNNNINGTKLNVAQMTGRDRGILKS 258  
 QY 162 QTIQOILDMGG-----GTWERDVTVRALRAAYNNPERAIDVLY-----SGIPEN 205  
 Db 259 EKVRHVVKIGSMHLKGVW--IPFERALAMAQR--EQIVDMLYPLFVRDIKRVIGTGVTPN 314  
 QY 206 VEAQFVARAPAGQQTNOQAASPAQPAVALPVQP-----SPASAGPHANPLNLPQGVPS 260  
 Db 315 AAA---ATAAAAATATASAPPPPPPPVAAAATTTAATAISKSSG-NGNSIS-----ATS 365  
 QY 261 GGSNPGVVPVPGAGSGALDALRQ-----LPQ---FQALLQLVQANPQIL-----Q 300  
 Db 366 GGSNVSGASGAGSTTSPVNTKAATAAGIPQNNYQYTNQ---QYTPQYGYQYNAPGNQNT 423  
 QY 301 PMLQELGKQNPQILRLIQENQAEFLRLVNESPEGG 335  
 Db 424 PASQPGSTTNDQYLQOQQQQQQQYQSNYYQGG 458

Search completed: December 17, 2003, 06:17:18  
 Job time : 15.5265 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:16:40 ; Search time 41.3907 Seconds  
(without alignments)  
2524.996 Million cell updates/sec

Title: US-09-805-550-2  
Perfect score: 2036  
Sequence: 1 MKNVTKLTGTFEIASPD.....BELTANYLLDRHGHEFDQQQ 405

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_23.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1211.5	59.5	389	10 Q9STA6	Q9sta6 lycopersico
2	1199.5	58.9	378	10 Q9FF16	Q9ff16 arabidopsis
3	1198.5	58.9	378	10 Q94C35	Q94c35 arabidopsis
4	1188.5	58.4	419	10 Q9M887	Q9m887 arabidopsis
5	1187.5	58.3	382	10 Q03990	Q03990 daucus caro
6	1127	55.4	392	10 Q40742	Q40742 oryza sativ
7	945.5	46.4	371	10 Q8LA46	Q8la46 arabidopsis
8	944.5	46.4	365	10 Q94CE9	Q94ce9 arabidopsis
9	925.5	45.5	379	10 Q03991	Q03991 daucus caro
10	840.5	41.3	367	10 Q9MA10	Q9ma10 arabidopsis
11	632	31.0	409	4 Q8WUB0	Q8wub0 homo sapien
12	576.5	28.3	362	11 Q8CAP3	Q8cap3 mus musculu
13	543	26.7	246	10 Q9S9L8	Q9s9l8 arabidopsis
14	501	24.6	341	5 Q9V135	Q9v135 dictyosteli
15	481	23.6	414	5 Q9V3W9	Q9v3w9 drosophila
16	478	23.5	414	5 Q9XZE0	Q9xze0 drosophila

17	436.5	21.4	372	5 Q23451	Q23451 caenorhabdi
18	416.5	20.5	296	11 Q8BRA9	Q8bra9 mus musculu
19	376.5	18.5	748	6 P79370	P79370 oryctolagus
20	369	18.1	389	5 Q81JS8	Q81js8 plasmodium
21	365.5	18.0	343	5 Q81MB7	Q81mb7 drosophila
22	361.5	17.8	113	10 Q9SA20	Q9sa20 arabidopsis
23	317	15.6	290	5 Q9VCD5	Q9vcd5 drosophila
24	271	13.3	65	10 Q9SCA8	Q9scas lycopersico
25	243	11.9	575	10 Q9FWF5	Q9fwf5 oryza sativ
26	231.5	11.4	551	10 Q94CS1	Q94cs1 arabidopsis
27	229.5	11.3	551	10 Q9SI18	Q9si18 arabidopsis
28	223	11.0	554	11 Q8C835	Q8c835 mus musculu
29	222	10.9	554	11 Q8K141	Q8k141 mus musculu
30	222	10.9	582	11 Q9JJ09	Q9jjp9 rattus norv
31	214.5	10.5	523	5 Q9NIF3	Q9nif3 dictyosteli
32	213	10.5	555	11 Q8C7T4	Q8c7t4 mus musculu
33	213	10.5	582	11 Q8R317	Q8r317 mus musculu
34	210	10.3	538	10 Q9SI19	Q9si19 arabidopsis
35	210	10.3	595	6 Q9SM59	Q9sm59 bos taurus
36	209	10.3	582	11 Q9QZM1	Q9qzm1 mus musculu
37	202.5	9.9	502	5 Q18672	Q18672 caenorhabdi
38	201.5	9.9	484	5 Q8MM76	Q8mm76 caenorhabdi
39	198	9.7	589	4 Q81XS9	Q81xs9 homo sapien
40	197.5	9.7	596	11 Q99NB8	Q99nb8 mus musculu
41	197	9.7	589	4 Q9UMX0	Q9umx0 homo sapien
42	196	9.6	601	4 Q9NR85	Q9nrs5 homo sapien
43	193	9.5	142	10 Q9LFI9	Q9lfi9 arabidopsis
44	193	9.5	454	5 Q8ML26	Q8ml26 caenorhabdi
45	193	9.5	589	4 Q9H0T8	Q9h0t8 homo sapien

## ALIGNMENTS

## RESULT 1

Q9STA6 PRELIMINARY; PRT; 389 AA.  
AC Q9STA6;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE RAD23 protein.  
GN RAD23.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.  
OX NCBI\_taxid=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. West virginia 106; TISSUE=fruit;  
RA Lemaire-Chamley M., Petit J., Raymond P., Chevalier C.;  
RT "Analysis of gene expression during early tomato fruit development by  
mRNA differential display."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ243875; CAB51544.1; .  
DR HSSP; P54725; 1DV0.  
DR InterPro; IPR004806; Rad23.  
DR InterPro; IPR006836; ST11.  
DR InterPro; IPR000449; UBA domain.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00627; UBA; 2.  
DR Pfam; PF00240; ubiquitin; 1.  
DR SMART; SM00727; ST11; 1.  
DR SMART; SM00165; UBA; 2.  
DR SMART; SM00213; UBQ; 1.  
DR TIGRfams; TIGR00601; rad23; 1.  
DR PROSITE; PS00053; UBIQUITIN 2; 1.  
SQ SEQUENCE 389 AA; 41508 MW; D2BAEDFOFE70778A CRC64;

Query Match 59.5%; Score 1211.5; DB 10; Length 389;  
Best Local Similarity 63.0%; Pred. NO. 1.3e-66;  
Matches 255; Conservative 43; Mismatches 88; Indels 19; Gaps 6;



```

DE Hypothetical 40.1 kDa protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Shinn P.,
RA Tracy S.E., Banh J., Bowser L., Carninci P., Chung M.K.,
RA Goldsmith A.D., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Quach H.L.,
RA Sakano H., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C.,
RA Toriumi M., Yamada K., Yu G., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037181; AAKS9766.1; -.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; ST11.
DR InterPro; IPR00449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00727; ST11; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBO; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS00553; UBIQUITIN_2; 1.
DR Hypothetical protein.
KW SEQUENCE 378 AA; 40075 MW; E49B591527A3EC27 CRC64;

Query Match      58.9%; Score 1198.5; DB 10; Length 378;
Best Local Similarity 62.9%; Pred. No. 86-66; 75; Indels 33; Gaps 9;
Matches 256; Conservative 43; Mismatches 43;

QY 1 MKNVKTILKGTNFETASPDASVADVKRIETTGQSTYRADOQMLIQGKILKDETTLE 60
DB 1 MKIFVKTILSGSNFEIVKPADKVDVKTAEIVKG-AYEPAKQMLIHQGVKLVKDETTLE 59

QY 61 SNGVAENSFLVIMLSKAKASSGASTAT-----TAKAPATLAQPAAPVAPAAASVARTPTQA 116
DB 60 ENNVVENSFVIMLSKTKASPSGASTAGAPASATQPTVATPQVS-APTASV----- 111

QY 117 PVATAETAPPSVQPOQAPATVAATDDADVYSQAASNLVFGNNLEQTICQILDGCGGTWE 176
DB 112 PVPTSGTA-----TAAAPA--TAASVQTDVYQQAASNLVAGTTLESTVQOILDGCGGSD 164

QY 177 RDTVVRLRAAYNNPERAIDYLSGIPENVEAQPVARAPAGQQTNOQAASPAQPAVALP 236
DB 165 RDTVVRLRAAFNNPERAVEVLYSGIPAQAEIPFVQAQAPATGEQANFLAQPOQAAA---- 221

QY 237 VQPSASAGPNANPLNLPQGVPSGGSGVGVGAGSGALDRLQLPQFALLQLVQANP 296
DB 222 --PAAATGCPNANPLNLPQGMFAADA-----GAGAGNLDFLRNHFQALRAMVQANP 273

QY 297 QILQPMQLBELGKQNPQILRLIQENQAEFLRLVNESPEGPGGNIILGQLAAAVPQLTVTP 356
DB 274 QILQPMQLBELGKQNPQLVRLIQHQADEFLRLINEPVEGEE--NVNQELEAAMPQAVTVTP 331

QY 357 EREAIORLEGMGFNRELVLVFFPACNCKDEELTANYLDHGHEPDDQ 403
DB 332 EREAIERLEGMGFDRAWLEVFFPACNCKEELANYLDHHPHEFDQ 378

RESULT 4
Q9M887 PRELIMINARY; PRT; 419 AA.
AC Q9M887;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)

```

```

DE Putative RAD23 (AT9g02540/F16B3_17).
GN F16B3.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.B., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F16B3 genomic sequence.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Gibson H.A., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,
RA Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Shinn P., Banh J., Bowser L.,
RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC021640; AAF32461.1; -.
DR EMBL; AY039562; AAK62617.1; -.
DR EMBL; AY113034; AAM47342.1; -.
DR HSSP; P54725; 1DV0.
DR InterPro; IPR002965; P rich extensn.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR PRINTS; PR01217; PRICHEXTENS.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS00553; UBIQUITIN_2; 1.
DR SEQUENCE 419 AA; 44247 MW; CAA13BC4FFEB1E25 CRC64;

Query Match      58.4%; Score 1188.5; DB 10; Length 419;
Best Local Similarity 58.4%; Pred. No. 3.7e-65;
Matches 253; Conservative 53; Mismatches 82; Indels 45; Gaps 10;

QY 1 MKNVKTILKGTNFETASPDASVADVKRIETTGQSTYRADOQMLIQGKILKDETTLE 60
DB 1 MKIFVKTILKGTNFETASPDASVADVKRIETTGQSTYRADOQMLIHQGVKLVKDETTIE 60

QY 61 SNGVAENSFLVIMLSKAKASSGASTAT--TAKA-----PATLAQPAAPVAP 105
DB 61 ENKVAENSFVIMNKSAPASAAASAGTSQAKSIPPTSPSTQPSISPTPASVAPVAP 120

QY 106 AAS----VARTPTQAPVATETAPPSVQPOQAPATVAAT-----DDADVYSQAAS 152
DB 121 APTRPPTPTPTAPVATETVTPPI-PEVPATISSSTPAPDSAPVSGDVGQAAS 179

QY 153 NLVFGNNLEQTICQILDGCGGTWERDVTVRLRAAYNNPERAIDYLSGIPENVEAQPA 212
DB 180 NLAAGSNLESTICQILDGCGGTWRETIVLALRAAFNNPERAVEVLYTGIPEQAEVPPVA 239

```

QY 213 RAPAAGQTNQQAASPAQPAVALPVQSPASAGPNANPLNLPQGVSPGSGSNPVGVPGAG 272  
Db 240 RPPASAGQPNPQAQQAASPAQPAVALPVQSPASAGPNANPLNLPQGVSPGSGSNPVGVPGAG 272  
QY 273 SGALDMLRQLPQFQALLQVQANPQILQPMQLQELGKQNPQILRLIQENQAELRLVNESP 332  
Db 287 AGTLDLRLNSQQFQALRAMVQANPQVQPMQLQELGKQNPQILRLIQENQAELRLVNESP 346  
QY 333 E-GGPGCNILGQAAAV--PQTLVTPERREATORLEGNGFNRELVLVFFACNKKDEILT 389  
Db 347 EGGESGNLIGQAAGMPQQAQVTHREERATERLEAMGFERLALVLEFFACNKKDEILT 406  
QY 390 ANYLLDHGHEFDD 402  
Db 407 ANYLLDHGHEFEE 419

RESULT 5  
Q03990 PRELIMINARY; PRT; 382 AA.  
ID O03990  
AC O03990  
DT 01-JUL-1997 (Tremblrel. 04, Created)  
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE RAD23, isoform I.  
OS Daucus carota (Carrot).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.  
OX NCBI\_TaxID=4039;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MO01C;  
RX MEDLINE=98345997; PubMed=9681019;  
RA Sturm A., Leinhardt S.;  
RT "Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in yeast."  
RL Plant J. 13:815-821(1998).  
DR EMBL; Y12013; CAA72741.1; --  
DR HSPSP; P54725; IDVO.  
DR InterPro; IPR002965; P-rich\_extensn.  
DR InterPro; IPR004806; Rad23.  
DR InterPro; IPR006636; STIL.  
DR InterPro; IPR000449; UBA domain.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00627; UBA; 2.  
DR SMART; SM00727; STIL; 1.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR SMART; SM00165; UBA; 2.  
DR SMART; SM00213; UBQ; 1.  
DR TIGRFAMs; TIGR00601; rad23; 1.  
DR PROSITE; PS50053; UBIQUITIN 2; 1.  
SQ SEQUENCE 382 AA; 40344 MW; 27FF02A2402F3DC2 CRC64;

Query Match 58.3%; Score 1187.5; DB 10; Length 382;  
Best Local Similarity 62.7%; Pred. No. 3.8e-65;  
Matches 255; Conservative 46; Mismatches 75; Indels 31; Gaps 11;

QY 1 MKNVTKLGTNFTEIASPDASVADVKRIIETTTQGGSTVRADQOMLIYQKILKDETTLE 60  
Db 1 MKIYVTKLGSQFEIQVNPDSVADVKRSIETAGAAVYPAQAOMLIYQKVLKDGTTLL 60

QY 61 SNGVAENSLVIMLSKAKASSSGASTATTA---KAPATLAQPAAPVAPAAASVARTTQAP 117  
Db 61 ENNVASNSFTVIMLSKSKSPSGSGSTTTSTAAPKAPQTSAPPSVP-APAVS-----QPP 113

QY 118 VATAET-AP-PSVQQAAPAAATVADDDVYSQAASNLVFGNNLEQTIQOILDMGGTTW 175  
Db 114 ASTLPVPAPSPAPATAPATIPSAVGSANVYDASAASLLVAGNLSEAIQOILDMGGTTW 173

QY 176 ERTLVVRLAAANNPERAIDLYSGIPENVEAQPVARAPAAAGQTNQQAASPA-OPAVA 234

Db 174 DRDTVIRVRAAFNNPERAVEILYSGIPEQAEAPPVAPSPSG-----QAANPLDQPPAA 228  
QY 235 LPVQSPASAGPNANPLNLPQGVSPGSGSNPVGVPGAGDALRLQLPQFQALLQVQAA 294  
Db 229 --AQAPASAGPNANPLNLPQGVSPGSGSNPVGVPGAGDALRLQLPQFQALLQVQAA 281  
QY 295 NPQILQPMQLQELGKQNPQILRLIQENQAELRLVNESPEGGPGGNILGQAAAVPQTULTV 354  
Db 282 NPQILQPMQLQELGKQNPQILRLIQENQAELRLVNESPEGGPGGNILGQAAAVPQTULTV 335

QY 355 TPERRAATORLEGNGFNRELVLVFFACNKKDEILTANYLLDHGHEFD 401  
Db 336 TPERRAATORLEGNGFNRELVLVFFACNKKDEILTANYLLDHGHEFE 382

RESULT 6  
Q40742 PRELIMINARY; PRT; 392 AA.  
ID Q40742  
AC Q40742  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE OBRAD23.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nipponbare;  
RX MEDLINE=97369378; PubMed=9225866;  
RA Schultz T.F., Quatrano R.S.;  
RT "Characterization and expression of a rice RAD23 gene."  
RL Plant Mol. Biol. 34:557-562(1997).  
DR EMBL; U63530; AAB65841.1; --  
DR HSPSP; P54725; IDVO.  
DR Gramene; Q40742; --  
DR InterPro; IPR004806; Rad23.  
DR InterPro; IPR006636; STIL.  
DR InterPro; IPR000449; UBA domain.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00627; UBA; 2.  
DR Pfam; PF00240; ubiquitin; 1.  
DR SMART; SM00727; STIL; 1.  
DR SMART; SM00165; UBA; 2.  
DR SMART; SM00213; UBQ; 1.  
DR TIGRFAMs; TIGR00601; rad23; 1.  
DR PROSITE; PS50053; UBIQUITIN 2; 1.  
SQ SEQUENCE 392 AA; 41754 MW; BD6E08574CC7CACB CRC64;

Query Match 55.4%; Score 1127; DB 10; Length 392;  
Best Local Similarity 59.1%; Pred. No. 2e-61;  
Matches 243; Conservative 48; Mismatches 90; Indels 30; Gaps 9;

QY 1 MKNVTKLGTNFTEIASPDASVADVKRIIETTTQGGSTVRADQOMLIYQKILKDETTLE 60  
Db 1 MKIYVTKLGSQFEIQVNPDSVADVKRSIETAGAAVYPAQAOMLIYQKVLKDGTTLL 60

QY 61 SNGVAENSLVIMLSKAKASSSGASTATTA---KAPATLAQPAAPVAPAAASVARTTQAP 118  
Db 61 ENNVASNSFTVIMLSKSKSPSGSGSTTTSTAAPKAPQTSAPPSVP-APAVS-----QAPV 111

QY 119 ATAETAP-----PSVQQAAPAAATVADDDVYSQAASNLVFGNNLEQTIQOILDMGG 173  
Db 112 APATTVPVTVSAPTATATSPAPAVASSEADNYGQATSNLVAGSNLEATIQSILEMGG 171

QY 174 TWRDVTVRALAAANNPERAIDLYSGIPENVEAQPVARAPAAAGQTN- QQAASPAQPA 232  
Db 172 IWRDVTVRALAAANNPERAIDLYSGIPENVEAQPVARAPAAAGQTN- PV--PPSPQANPTQASQATQA 228

QY 233 VALPVQSPASAGPNANPLNLPQGVSPGSGSNPVGVPGAGDALRLQLPQFQALLQV 292



Db 229 A-----PSILSGPNASFLDLPQALPNASTD-----AAGLNLDAALRNNAQFRTLLSLV 278  
Qy 293 QANPQILQMLQELGKQNPQILRLIQENQABFLRLVNESPEGGPG-GNILGQLAAAVPQTL 352  
D 279 QANPQILQMLQELGKQNPQILRLIQENQABFLRLVNESPEGGPG-GNILGQLAAAVPQTL 338  
Qy 353 TVTPPERAIQRLGCMGFNRELVLVFFACNKKDELTANYLLDHGHEFDDQ 403  
D 339 AVTPPEDEAILRLGPMGFDRALVLDVFFACNKKDELTANYLLDHGHEFDDQ 389

RESULT 7  
Q8LA46 PRELIMINARY; PRT; 371 AA.  
ID Q8LA46  
AC Q8LA46  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE DNA repair protein RAD23, putative.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC euroside II; Brassicales; Brassicaceae; Arabidopses.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,  
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length messenger RNA sequences greatly improve genome  
RT annotation.";  
RL Genome Biol. 0:0-0 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.;  
RT "Full-length cDNA from Arabidopsis thaliana.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY088037; AAM65583.1; -  
DR InterPro; IPR004806; Rad23.  
DR InterPro; IPR006636; STI1.  
DR InterPro; IPR000449; UBA domain.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00627; UBA; 2.  
DR SMART; SM00727; STI1; 1.  
DR SMART; SM00165; UBA; 2.  
DR SMART; SM00213; UBQ; 1.  
DR TIGRfams; TIGR00601; rad23; 1.  
DR PROSITE; PS0053; UBIQUITIN 2; 1.  
SQ SEQUENCE 371 AA; 39747 MW; 081493086E9A976E7 CRC64;

Query Match 46.4%; Score 945.5; DB 10; Length 371;  
Best Local Similarity 50.7%; Pred.No.2.5e-50;  
Matches 208; Conservative 60; Mismatches 95; Indels 47; Gaps 11;  
SQ SEQUENCE 371 AA; 39747 MW; 081493086E9A976E7 CRC64;

Qy 1 MKNVKTILKGTNFEIASPASVADVKRIIETQGSTYRADQMLIYQGIKDETTLE 60  
D 1 MKNVKTILKGTNFEIASPASVADVKRIIETQGSTYRADQMLIYQGIKDETTLE 60  
Qy 61 SNGVAENSLVIMLSKAKA-SGASSTATTAKAPATLAQPAAPVAPASVARTPTQAPVAT 120  
D 61 ENKVTTEGFLVVLMSKSK---SGSAGASVQTSVSPV-----SATTSTTKPAAP-ST 111  
Qy 121 AETAPSVOP---QAAPATVAATDDADVYSQASNLVFGNNLEQIQLDMGGGTWER 177  
D 121 TQSSVPVAPSPIPAQSQPAA-----QTDYTGAAATLVSGSLSEQVMQVQIMEMGGGWDK 165  
Qy 178 DTWVTRALRAYNNPERAIDLYSGIPENVE-AQPVARAPAAQQTNOQAASPAQVALP 236  
D 166 EIVTRALRAYNNPERAIDLYSGIPQTAEVAVPPEAQIAG-----SGAAP 212  
Qy 237 VQPSFASAGPNANPLNLPQGVPSGSGNPGVVPVAGS---GALDALRLQLPQALLQLVQ 293  
D 213 V--APASGPGNSPLDLPQ-----ETVAAGSGDLGLTFLEFLRNNDQFQQLRTWVH 261

Qy 294 ANPQILQMLQELGKQNPQILRLIQENQABFLRLVNESPEGGPG-GNILGQLAAAVPQTL 352  
D 262 SNPQILQMLQELGKQNPQILRLIQENQABFLRLVNESPEGGPG-GNILGQLAAAVPQTL 321  
Qy 353 TVTPPERAIQRLGCMGFNRELVLVFFACNKKDELTANYLLDHGHEFDD 402  
D 322 NVTPAEQAIRLEAMGFDRALVTEAFACNKKDELTANYLLDHGHEFDD 371

RESULT 8  
Q94CE9 PRELIMINARY; PRT; 365 AA.  
ID Q94CE9  
AC Q94CE9  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Putative RAD23 protein.  
GN F20B17.8.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC euroside II; Brassicales; Brassicaceae; Arabidopses.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,  
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayaishizaki Y., Ishida J.,  
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,  
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,  
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,  
RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;  
RT "Full length cDNA of gene F20B17.8 (GI:7715605).";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,  
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,  
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
RA Hayaishizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,  
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,  
RA Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Full length cDNA of gene F20B17.8 (GI:7715605).";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY034912; AAK59419.1; -  
DR EMBL; AY063103; AAL34277.1; -  
DR InterPro; IPR004806; Rad23.  
DR InterPro; IPR006636; STI1.  
DR InterPro; IPR000449; UBA domain.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00627; UBA; 2.  
DR Pfam; PF00240; ubiquitin; 1.  
DR SMART; SM00727; STI1; 1.  
DR SMART; SM00165; UBA; 2.  
DR SMART; SM00213; UBQ; 1.  
DR TIGRfams; TIGR00601; rad23; 1.  
DR PROSITE; PS0053; UBIQUITIN 2; 1.  
SQ SEQUENCE 365 AA; 39157 MW; 4FED9EC59B467745 CRC64;

Query Match 46.4%; Score 944.5; DB 10; Length 365;  
Best Local Similarity 50.5%; Pred.No.2.8e-50;  
Matches 206; Conservative 56; Mismatches 97; Indels 49; Gaps 9;  
SQ SEQUENCE 365 AA; 39157 MW; 4FED9EC59B467745 CRC64;

Qy 1 MKNVKTILKGTNFEIASPASVADVKRIIETQGSTYRADQMLIYQGIKDETTLE 60  
D 1 MKNVKTILKGTNFEIASPASVADVKRIIETQGSTYRADQMLIYQGIKDETTLE 60  
Qy 61 SNGVAENSLVIMLSKAKA-SGASSTATTAKAPATLAQPAAPVAPASVARTPTQAPVAT 119  
D 61 ENKVTTEGFLVVLMSKSKSGSAGASVQTSVSPV-----SATTSTTKPAAP-----STTQSSPV 111



```

QY 120 TAETAPPSVQPAAPAAATVADDDADVYSQAASNLVFGNNLEQTIQOILDMGGGTWERDT 179
DB 112 PASPIPAEQPAA-----QTDYTGAASTLVSGSLEQMVQVQIMEGGSGWDXET 161
QY 180 VTRALRAAYNNPERAIDYLSGIPENVE-AQVARAPAAAGQQTNOQAASPAQPAVALPVQ 238
DB 162 VTRALRAAYNNPERAIDYLSGIPQTAQAEVAVPVPEAQIAG-----SGAAPV- 207
QY 239 PSASAGPNANPLNLPQGVPSGSGNPGVVPVPGAGS---GALDALRQLPQFQALLQLOVAN 295
DB 208 -APASGPNNSPLDLFPFQ-----ETVAAGSGDLGTLFLRNNDQFQQLRTVHSHN 257
QY 296 PQILQPMQLGKQNPQILRLIQENQAFLRLVNESPEGPGF-GNIIQLQAAAAVPOTLTV 354
DB 258 PQILQPMQLGKQNPQILRLIQENQAFLRLVNEPYEGSGDGEDMFDQEQEMPHAINV 317
QY 355 TPEREAIQRLGEGFNRRLVLEVFACNCKDEELTANYLLDHGHEFDD 402
DB 318 TPAEQEAIQRLGEGFNRRLVLEVFACNCKDEELTANYLLDHGHEFDD 365

RESULT 9
O03991
ID O03991 PRELIMINARY; PRT; 379 AA.
AC O03991;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RAD23 protein, isoform II.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W001C;
RX MEDLINE=98345997; PubMed=9681019;
RA Sturm A., Leinhard S.;
RT "Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in yeast.";
RL Plant J. 13:815-821(1998).
DR EMBL; Y12014; CAA72742.1; -.
DR HSP; P54725; IDV0.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STII.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00727; STII; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
SQ SEQUENCE 379 AA; 40530 MW; B26697B439CC5929 CRC64;

Query Match
Best Local Similarity 50.2%; Score 925.5; DB 10; Length 379;
Matches 203; Conservative 57; Mismatches 117; Indels 27; Gaps 8;

QY 1 MKNLVKLTGNTFIEASPDASVADVKRIETTGQSTYRADQOQLYQGIKDETTLE 60
DB 1 MKLTVKLTGSHFFIRAPNDTVMKNIIEDLQKDNYPGCGQLLIHNGKVLKDESTLA 60
QY 61 SNGVAENFLVIMLSKAKA-SSSGASTATTAKAPATLAQAPVAPAAASVARTPTQAPVA 119
DB 61 ESKISEGFLVVMLGKSTNSSTGTPAAQSSASAPTPAPAVAPAPAAAPASAVIPNT 120
QY 120 TAETAPPSVQPAAPAAATVADDDADVYSQAASNLVFGNNLEQTIQOILDMGGGTWERDT 179
DB 121 TVPEAP--LSPAPAP-----SDTYGEAASNVVAGSNLEQTIQIHMDGGGMDTNN 169

```

```

QY 180 VVRALRAAYNNPERAIDYLSGIPENVEAQ-PVARAPAAAGQQTNOQAASPAQPAVALPVQ 238
DB 170 VSRALRAAYNNPERAIDYLSGIPMAEAAVPVSH--FQDQINAGNNAISDNGVA---- 223
QY 239 PSASAGPNANPLNLPQGVPSGSGNPGVVPVPGAGSGALDALRQLPQFQALLQLOVANPOI 298
DB 224 -GAAPGAPNSLPLNFPQBTLSG-----VTGAGLGSLEFLRNPFQFQLRSMVORNPQI 276
QY 299 LQPMQLGKQNPQILRLIQENQAFLRLVNESPEGPGGNILGOLAAAAVPOTLTVTPPE 358
DB 277 LQPMLELQKQNPQILRLIQEHHEFLQLINEPVEASE-GDMFDQEQDVPQBITVTAAD 335
QY 359 REAIQRLGEGFNRRLVLEVFACNCKDEELTANYLLDHGHEFDD 402
DB 336 QEAIRLEAMGDFRGLVIEAFACDRNEELAVNYLLENAGDFED 379

RESULT 10
Q9MA10
ID Q9MA10 PRELIMINARY; PRT; 367 AA.
AC Q9MA10;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE F20B17.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharshy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F20B17 from chromosome 1."
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howng B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharshy N., Southwick A.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC010793; AAF68123.1; -.
DR HSP; P54725; IDV0.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
SQ SEQUENCE 367 AA; 39790 MW; C6D01134B5C69CE6 CRC64;

```

```

Query Match      41.3%; Score 840.5; DB 10; Length 367;
Best Local Similarity 47.4%; Pred. No. 6.7e-44;
Matches 192; Conservative 53; Mismatches 119; Indels 41; Gaps 8;

QY 1 MKNVKTLLKGNTEFEASPDASVADVVKRIETTGQSTYRADQOQMLIYOGKILKDETTLE 60
DB 1 MKNVKTLLKGNTEFEIRVLPSTIMAVKKNIEDSQSKDNYPCQQLLIHNGKVLKDETSLV 60

QY 61 SNGVAENSFLVIMLSKAKASSSGASTATTAKAPATAQAAPVAPAAASV 109
DB 61 EYKIDKFNFFVVMVTKKAVSTPAPATTQOSAPASTTAVTSSSTTTTVAQAFTPV-PALAP 119

QY 110 ARTPTQ-----APVATAETAPPSVQOQAPAAAT-VAATDD-----ADVYSQ 149
DB 120 TSTPASTPASATASSEPAASAAKQEPKPAETPVATSPATYDSTSGDSSSNLFPED 179

QY 150 AASNLVFGNNLEOTIQIILDMGGGTWERTVVRALRAAYNNPERAIDLYSGIPENVEAQ 209
DB 180 ATSLVVTGQSYENMVTEIMSGM-----YERQVIAALRASFNPNPDRAVEYLLWGI PGDRSQ 236

QY 210 PVARAPAAQOQTNQOAS---PAQPAVALPVQPSPASAGNPANPLNFPQGVPGSGSNP 266
DB 237 AVDDPP-----QAASTGVPOSSAVAAAAATTTATT-----TTTSSGGHP- 275

QY 267 VVPGAGSGALDRLQFQALLQLOVANPOILOQMLQELGKONPOILRLIQENQABFLR 326
DB 276 -----LEPLRNQFQFQMRQIIQONPSLLPALLQOIGRENPOQLLIQSQHEFTIQ 326

QY 327 LVNE-----SPEGPGGNILGQLAAAVP---QTLVTTPPEERBAIORLEGMGFNRELVLV 378
DB 327 MLNEPVQEGAGQGGGGGGGGIAEAGSGHNMVYIQTVPQEKAIERLKAIGFPEGLVIA 386

QY 379 FPACNKBELTANYLLDHGHEFD 401
DB 387 YFACENKLENLAANFLQONFED 409

RESULT 12
Q8CAP3 PRELIMINARY; PRT; 362 AA.
AC Q8CAP3;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE RAD23a homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK038300; BAC29962.1; -.
SQ SEQUENCE 362 AA; 39635 MW; 804B2608ECA241A9 CRC64;

Query Match      28.3%; Score 576.5; DB 11; Length 362;
Best Local Similarity 33.0%; Pred. No. 9.8e-28;
Matches 138; Conservative 80; Mismatches 123; Indels 77; Gaps 13;

QY 3 LNVKTLKGNTEFEIASPDASVADVVKRIETTGQSTYRADQOQMLIYOGKILKDETTLESN 62
DB 5 ITLTLQOQTFKIRMEPDETIVKLEKIEAEKRDAPFVAGQKLIYAGKILKDDVPIKBY 64

QY 63 GVAENSFLVIMLSKAKASSSGASTATTAKAPATAQAAPVAP--AASVARTP----- 113
DB 65 HIDEKNFVVMVTKAKAQO--GIPAPPEASTAVPEPSTPPFVLAGSMSPPTSRDK 122

QY 114 TQAPVATAETAPPSVQOQAPAAATVAATDADVYSQAASNLVFGNNLEQTTIQIILDMGGG 173
DB 123 SPSESTTTTSPESIGSVPSGSGSGREED-----AASLTVTGSEVETMLTEIMSGM-- 174

QY 174 TWEDTIVVRALRAAYNNPERAIDLYSGIPENVEAQPVARAPAAQOQTNQOASPAQAV 233
DB 175 -YERERVAALRASNNPHRAVEYLLTGIGSPSE-----PEHG--SVQESQAEQPAT 224

Query Match      41.3%; Score 840.5; DB 10; Length 367;
Best Local Similarity 47.4%; Pred. No. 6.7e-44;
Matches 192; Conservative 53; Mismatches 119; Indels 41; Gaps 8;

QY 1 MKNVKTLLKGNTEFEASPDASVADVVKRIETTGQSTYRADQOQMLIYOGKILKDETTLE 60
DB 1 MKNVKTLLKGNTEFEIRVLPSTIMAVKKNIEDSQSKDNYPCQQLLIHNGKVLKDETSLV 60

QY 61 SNGVAENSFLVIMLSKAKASSSGASTATTAKAPATAQAAPVAPAAASV 109
DB 61 EYKIDKFNFFVVMVTKKAVSTPAPATTQOSAPASTTAVTSSSTTTTVAQAFTPV-PALAP 119

QY 110 ARTPTQ-----APVATAETAPPSVQOQAPAAAT-VAATDD-----ADVYSQ 149
DB 120 TSTPASTPASATASSEPAASAAKQEPKPAETPVATSPATYDSTSGDSSSNLFPED 179

QY 150 AASNLVFGNNLEOTIQIILDMGGGTWERTVVRALRAAYNNPERAIDLYSGIPENVEAQ 209
DB 180 ATSLVVTGQSYENMVTEIMSGM-----YERQVIAALRASFNPNPDRAVEYLLWGI PGDRSQ 236

QY 210 PVARAPAAQOQTNQOAS---PAQPAVALPVQPSPASAGNPANPLNFPQGVPGSGSNP 266
DB 237 AVDDPP-----QAASTGVPOSSAVAAAAATTTATT-----TTTSSGGHP- 275

QY 267 VVPGAGSGALDRLQFQALLQLOVANPOILOQMLQELGKONPOILRLIQENQABFLR 326
DB 276 -----LEPLRNQFQFQMRQIIQONPSLLPALLQOIGRENPOQLLIQSQHEFTIQ 326

QY 327 LVNE-----SPEGPGGNILGQLAAAVP---QTLVTTPPEERBAIORLEGMGFNRELVLV 378
DB 327 MLNEPVQEGAGQGGGGGGGGIAEAGSGHNMVYIQTVPQEKAIERLKAIGFPEGLVIA 386

QY 379 FPACNKBELTANYLLDHGHEFD 401
DB 387 YFACENKLENLAANFLQONFED 409

RESULT 11
Q8WUB0 PRELIMINARY; PRT; 409 AA.
AC Q8WUB0;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE RAD23 homolog B (S. cerevisiae).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020973; AAH20973.1; -.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006836; STII.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00727; STII; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS0053; UBIQUITIN 2; 1.
SQ SEQUENCE 409 AA; 43199 MW; 475FBD499DACAC69 CRC64;

Query Match      31.0%; Score 632; DB 4; Length 409;
Best Local Similarity 34.5%; Pred. No. 4.5e-31;
Matches 153; Conservative 81; Mismatches 133; Indels 76; Gaps 12;

QY 1 MKNVKTLLKGNTEFEASPDASVADVVKRIETTGQSTYRADQOQMLIYOGKILKDETTLE 60
DB 1 MKNVKTLLKGNTEFEIRVLPSTIMAVKKNIEDSQSKDNYPCQQLLIHNGKVLKDETSLV 60

```

```
QY 234 ALPVQSPASAGNPANPLNLFPGQVPSGSGNSPGVPGAGSGALDRLQLPQFQALLQLVQ 293
Db 225 -----EAGENP-----LEFLRQPFQFQFQFQFQFQFQFQFQFQFQFQFQFQFQ 248
QY 294 ANPQILQPMQLQELGKQNPQLRLIQENQAEFLVLVNESPGGPGGNI-----LQQLAAA 347
Db 249 QNPALLPALLQQLQGOENPQQLQISRHQEQFIQMLNEPP--GELADISDVEGEVGAIGSE 306
QY 348 VPQT--LTVTPBERIAQRLGEGFNRELVLVFFACNKKDELTANYLLDHGHEFDQ 403
Db 307 APQMNVIQVTPQEKAIERLKGFPESLVIQAFACEKNENLAANFL--SQNFDE 362

RESULT 13
Q9S9L8 PRELIMINARY; PRT; 246 AA.
AC Q9S9L8;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE T24D18.27 protein.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Liu S.X., Yu G., Sakano H., Jhaveri A., Lee J.M., Lenz C., Pham P.,
RA Toriumi M., Chin C., Chio J., Choi E., Chung M., Gonzalez A.,
RA Howing B., Koo T., Li J., Liu A., Vaysberg M., Altafi H., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A., Hansen N., Johnson-Hopson C.,
RA Khan S., Kim C., Lam B., Nguyen M., Palm C., Shinn P., Tambunga G.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN 2;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010924; AAF18513.1; -.
DR HSSP; P02248; IUBI.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 1.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00165; UBA; 1.
DR SMART; SM00213; UBQ; 1.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
SQ SEQUENCE 246 AA; 25966 MW; 0867D1E7D72B6FFE CRC64;

Query Match 26.7%; Score 543; DB 10; Length 246;
Best Local Similarity 41.8%; Pred. No. 6.9e-26;
Matches 117; Conservative 41; Mismatches 74; Indels 48; Gaps 4;

QY 1 MKNLVKTLKGTNFEIEASPDASVADVKRIIETQGSTYRADQQLIYQKILKDETTLE 60
Db 1 MKNLVKTLKGSHEIRVLPDTTMAVKKNIEDSQSDKNYPCGQQLIHNGKVLKDETTLV 60
QY 61 SNGVAENSLVIMLSKAKASSGASTATTAKAPATLAQPAAPVAPASVARTTQAPVAT 120
Db 61 ENKVTGFLVLMLSKSKTASSAGSPSTQLAARSTTQSIAPASNSTPVEQFTA----- 115
QY 121 AETAPPSVQPAAPATAATDADVYSQAASNLVFNQNLLEQTIQILDGCGGTWERDVT 180
Db 116 -----QSDTYGQAASLTLSVSSSTEQMVQIMMGSGSWDKETV 153
QY 181 VRLAAYNNPRAIDYLSGIPENVEAQVAPARAAGQTNQQAASPAQPAVALPVQPS 240
Db 154 TRALRAYNNPRAVDYLSGIPETV---TIPATNLSGVSGSRELTAP----- 198
```

```
QY 241 PASAGNPANPLNLFPGQVPSGSGNSPGVPGAGSGALDRL 280
Db 199 PPSGGPNSPLDFPQEAVSAA-----GGDLGTLEFLR 232

RESULT 14
O97135 PRELIMINARY; PRT; 341 AA.
AC O97135;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE RepC-binding protein A.
GN RCBA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=slb153;
RA Li G., Alexander H., Alexander S.;
RT "rcba, the Dictyostelium discoideum homolog of yeast repair gene RAD23.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF103870; AAD17913.1; -.
DR HSSP; P54725; IDVO.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STI1.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00727; STI1; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
SQ SEQUENCE 341 AA; 37528 MW; E376B909B6DEB57E CRC64;
```

```
Query Match 24.6%; Score 501; DB 5; Length 341;
Best Local Similarity 31.9%; Pred. No. 3.9e-23;
Matches 136; Conservative 59; Mismatches 123; Indels 108; Gaps 15;

QY 1 MKNLVKTLKGTNFEIEASPDASVADVKRIIETQGSTYRADQQLIYQKILKDETTLE 60
Db 1 MKVTIKNKKEIYVFEVNGDLTVAEKKNLISEKHQT---PSWQTLIYSGKILEDKRTLE 57
QY 61 SNGVAENSLVIMLSKAKASSGASTATTAKAPATLAQPAAPVAPASV----- 109
Db 58 SYNITDSGFIKXMKPR-----EAPAT--TPAPSTPHLNYISTNHHYCR 103
QY 110 -----ARTPTQAPVATAETAPSPVQQAAPATAATDADVYSQAASNLVFN 158
Db 104 NNHKLNTNTSTTTPTPTNT--PATPNPTPTTSTPGSTSTTS--POOSSOFATGT 160
QY 159 NLEQTIQILDGCGGTWERDVTVRLAAYNNPRAIDYLSGIPENVEAQVAPARAAG 218
Db 161 ELEATIKNTDNG---FARDQVLRALRTLTFNNAEIAEVLVSG-----NIPAA 206
QY 219 QQTNOQAASPAQPAVALPVQSPASAGNPANPLNLFPGQVPSGSGNSPGVPGAGSALDA 278
Db 207 DPEDEEE-----MEGGG-----GSGDNPF 227
QY 279 LRQLPQFQALLQLVQNPQLQPMQLQELGKQNPQLRLIQENQAEFLVLVNESPGGPG 338
Db 228 LRNHFNRLRAISKNPSIIFGILQQLAQTNPALVRQIQENPNFIRLF--QGDNPG 285
QY 339 NILGOLAAVPTLTTPEREAIQRLGEGFNRELVLVFFACNKKDELTANYLLDHG 397
Db 286 N-PGQF-----TLQVTQESEAIRLQALTGMDKSTVIEAFACDKNEELTASYLFE 338
QY 398 HEFDQ 403
```

```

Db          |||
339 ---DDE 341

RESULT 15
Q9V349      PRELIMINARY; PRT; 414 AA.

AC Q9V3W9;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE RAD23 protein.
GN RAD23 ORF C61836.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jafarizadeh S., Kalish J., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[2]
RN SEQUENCE FROM N.A.
RP Brodey M.H., Rubin G.M., Tsang G.;
RT "Full Length Drosophila melanogaster cDNA sequence.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003844; AAF59352.1; -
DR EMBL; AF132147; AAD33594.1; -
DR HSP; P54725; IDV0.
DR FlyBase; FBgn026777; Rad23.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; ST11.
DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR000626; Ubiquitin.

```

---

```

DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00727; ST11; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS00553; UBIQUITIN_2; 1.
SQ SEQUENCE 414 AA; 45780 MW; 4C2E494CA116F7AB CRC64;

Query Match      23.6%; Score 481; DB 5; Length 414;
Best Local Similarity 29.3%; Pred. No. 8.2e-22;
Matches 134; Conservative 72; Mismatches 153; Indels 98; Gaps 10;

Qy 1 M K L N V K T L K G T N F E I E A S P D A S V A D V K R I I E T T Q G Q S T Y R A D Q Q M L I Y Q G K I L K D E T T L E 60
Db 1 M I I T I K N L Q Q T F T I E P A P E K T V L E L K K I F E E R G P E - Y V A E K Q K L I Y A G V I L T D D R T V G 59
Qy 61 S N G V A E N S F L V I M L S K A K A S S S -----G A S T A T T A K A P A T L A Q P A A P A P A A S V A R T P T 114
Db 60 S Y N V D E K K F T V V M L T R D S S S S N R N Q L S V K E S N K L T S T D D S K Q S M P C E A N H T N S P S S T N T 119
Qy 115 Q A P V A T A E T A P P S V Q P Q A A P A A T V A A T D D A D V Y S Q A A S N L V F G N N L E S Q T I Q Q I L D M G G T 174
Db 120 E D S V L S R E T R P L S D E L I G E L A -----Q A S L Q S R A E S N L M G D S Y N Q T V L S M V E M G --- 170
Qy 175 W E R D T V V R A L R A A Y N N P E R A I D Y L Y G I P E N V E A Q P V A R A P A A G O O T N Q Q A A S P A Q P A V A 234
Db 171 Y P R E Q V E R A M A A S Y N N P E R A V E L I N G I P -----A E E G T F Y N R L N E S T P S L 217
Qy 235 L P V Q P S P A S A G P N A P N L N L P F Q G V P S G S N P G V P G A G S G A L D A L R O L P Q F A L L Q L V Q A 294
Db 218 I P S G P Q P A S A -----T S A E R S T E S N -----S D P F B F L R S Q P Q F L Q M R S L I Y Q 259
Qy 295 N P Q I L Q P M L Q E L G K N Q P Q I L R L I Q E N Q A E F L R L V N E --S P E G G P G G N I L Q L A A A V P Q T L 352
Db 260 N P H L L H A V L Q Q I G Q T N P A L L Q L I S E N Q D A F I N L M N Q P I D R E S E S G A T V P P V S N A R I P S T L 319
Qy 353 -----T V T P E E R E A I Q R L 365
Db 320 D N V D L F S P D L E V A T S A Q R S A A C T S A A H Q S G S A A D N E D L E Q P L G V S T I R L N R Q D K A I E R L 379
Qy 366 E G M G F N R E L V L E V F F A C N K D E E L T A N Y L L D H G H E F D D 402
Db 380 K A L G F P E A L V L Q A Y F A C E K N E Q A A N F L L --S S S F D D 414

Search completed: December 17, 2003, 06:20:06
Job time : 45.3907 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:16:39 ; Search time 38.771 Seconds  
(without alignments)  
1658.049 Million cell updates/sec

Title: US-09-805-550-2

Perfect score: 2036

Sequence: 1 MKNVKTLLKGTNFEIASEPD.....EELTANYLLDHGHEFDQDQ 405

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2036	100.0	405	21	Maize Rad23 protei
2	1199.5	58.9	378	21	Arabidopsis thalia
3	1199.5	58.9	378	21	Arabidopsis thalia
4	1095	53.8	335	21	Arabidopsis thalia
5	1095	53.8	335	21	Arabidopsis thalia
6	981	48.2	307	21	Arabidopsis thalia
7	981	48.2	307	21	Arabidopsis thalia
8	945.5	46.4	371	21	Arabidopsis thalia
9	913	44.8	368	21	Arabidopsis thalia

10	900.5	44.2	368	21	AAV71459	Maize Rad23 protei
11	881.5	43.3	348	21	AAG17378	Arabidopsis thalia
12	850	41.7	345	21	AAG44342	Arabidopsis thalia
13	754.5	37.1	299	21	AAG17379	Arabidopsis thalia
14	732	36.0	257	21	AAG36529	Arabidopsis thalia
15	725	35.6	296	21	AAG44343	Arabidopsis thalia
16	634	31.1	409	19	AAW75700	Vpr protein bindin
17	634	31.1	409	19	AAW68186	Vpr protein bindin
18	632	31.0	409	24	ABU07460	Protein different
19	628.5	30.9	214	21	AAG36530	Arabidopsis thalia
20	623.5	30.6	416	23	ABB57171	Mouse ischaemic co
21	573	28.1	363	18	AAW23658	E6AP-binding prote
22	573	28.1	363	19	AAW75699	Vpr protein bindin
23	573	28.1	363	19	AAW68185	Vpr protein bindin
24	573	28.1	375	24	ABU11879	Human ABCAL intera
25	572.5	28.1	379	21	AAW58841	Breast and ovarian
26	548	26.9	346	18	AAW21730	Nuclear mitotic ap
27	514.5	25.3	186	21	AAG36531	Arabidopsis thalia
28	481	23.6	414	22	ABB58584	Drosophila melanog
29	347	17.0	179	19	AAW75843	Vpr protein bindin
30	347	17.0	179	19	AAW68187	Vpr protein bindin
31	324	15.9	174	19	AAW75850	Protein sequence o
32	324	15.9	174	19	AAW68194	Protein sequence o
33	318	15.6	117	23	ABP34618	Human ORF3591 prot
34	317	15.6	290	22	ABP72011	Drosophila melanog
35	307	15.1	136	21	AAG33942	Arabidopsis thalia
36	289	14.2	150	19	AAW75849	Protein sequence o
37	289	14.2	150	19	AAW75848	Protein sequence o
38	229	11.2	112	19	AAW75848	Protein sequence o
39	229	11.2	112	19	AAW68192	Protein sequence o
40	211	10.4	536	21	AAG30040	Arabidopsis thalia
41	211	10.4	577	21	AAG30039	Arabidopsis thalia
42	201	9.9	604	22	ABG16557	Novel human diagno
43	197.5	9.7	596	22	AAG64210	Murine HSP47 inter
44	197	9.7	589	22	AAW94008	Human stomach canc
45	197	9.7	589	22	AAB94311	Human protein sequ

ALIGNMENTS

RESULT 1

AAV71458  
ID AAV71458 standard; Protein; 405 AA.

AC AAV71458;

DT 04-OCT-2000 (first entry)

DE Maize Rad23 protein #1.

KW Rad23; maize; ATCC No: PTA-530; recombinant expression cassette; wheat;  
transgenic plant; soybean; sunflower; sorghum; canola; modulator.

OS Zea mays.

PN WO200031268-A1.

PD 02-JUN-2000.

PF 12-OCT-1999; 99WO-US24129.

PR 23-NOV-1998; 98US-0109728.

PA (PTON-) PIONEER HI-BRED INT INC.

PI Mahajan PB, Tagliani L;

DR WPI; 2000-400078/34.

XX N-PSDB; AAD01230.

PT Isolated nucleic acid encoding maize RAD23 protein is used to modulate  
the levels of polypeptides in plant or in assays for identifying

PR	25-FEB-1999	99US-0121825
PR	05-MAR-1999	99US-0123180
PR	09-MAR-1999	99US-0123548
PR	23-MAR-1999	99US-0125788
PR	25-MAR-1999	99US-0126264
PR	29-MAR-1999	99US-0126785
PR	01-APR-1999	99US-0127432
PR	06-APR-1999	99US-0128234
PR	08-APR-1999	99US-0128714
PR	16-APR-1999	99US-0129845
PR	19-APR-1999	99US-0130077
PR	21-APR-1999	99US-0130449
PR	23-APR-1999	99US-0130510
PR	28-APR-1999	99US-0130891
PR	28-APR-1999	99US-0131449
PR	30-APR-1999	99US-0132048
PR	30-APR-1999	99US-0132407
PR	04-MAY-1999	99US-0132484
PR	05-MAY-1999	99US-0132485
PR	06-MAY-1999	99US-0132486
PR	06-MAY-1999	99US-0132487
PR	07-MAY-1999	99US-0132863
PR	11-MAY-1999	99US-0134256
PR	14-MAY-1999	99US-0134218
PR	14-MAY-1999	99US-0134219
PR	14-MAY-1999	99US-0134221
PR	14-MAY-1999	99US-0134370
PR	18-MAY-1999	99US-0134768
PR	19-MAY-1999	99US-0134941
PR	20-MAY-1999	99US-0135124
PR	21-MAY-1999	99US-0135353
PR	24-MAY-1999	99US-0135629
PR	25-MAY-1999	99US-0136021
PR	27-MAY-1999	99US-0136392
PR	28-MAY-1999	99US-0136782
PR	01-JUN-1999	99US-0137222
PR	03-JUN-1999	99US-0137528
PR	04-JUN-1999	99US-0137724
PR	07-JUN-1999	99US-0137724
PR	08-JUN-1999	99US-0138094
PR	10-JUN-1999	99US-0138540
PR	10-JUN-1999	99US-0138847
PR	14-JUN-1999	99US-0139119
PR	16-JUN-1999	99US-0139452
PR	16-JUN-1999	99US-0139453
PR	17-JUN-1999	99US-0139492
PR	18-JUN-1999	99US-0139454
PR	18-JUN-1999	99US-0139455
PR	18-JUN-1999	99US-0139456
PR	18-JUN-1999	99US-0139457
PR	18-JUN-1999	99US-0139458
PR	18-JUN-1999	99US-0139459
PR	18-JUN-1999	99US-0139460
PR	18-JUN-1999	99US-0139461
PR	18-JUN-1999	99US-0139462
PR	18-JUN-1999	99US-0139463
PR	18-JUN-1999	99US-0139750
PR	18-JUN-1999	99US-0139763
PR	21-JUN-1999	99US-0139817
PR	22-JUN-1999	99US-0139899
PR	23-JUN-1999	99US-0140353
PR	23-JUN-1999	99US-0140354
PR	24-JUN-1999	99US-0140695
PR	28-JUN-1999	99US-0140823
PR	29-JUN-1999	99US-0140891
PR	30-JUN-1999	99US-0141287
PR	01-JUL-1999	99US-0141842
PR	01-JUL-1999	99US-0142154
PR	02-JUL-1999	99US-0142055
PR	06-JUL-1999	99US-0142390
PR	08-JUL-1999	99US-0142803
PR	09-JUL-1999	99US-0142920
PR	12-JUL-1999	99US-0142977

PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 17-JUL-1999; 99US-0144086.  
PR 18-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145193.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0155659.  
PR 29-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 58.9%; Score 1199.5; DB 21; Length 378;  
Best Local Similarity 62.9%; Pred. No. 5.8e-78;  
Matches 256; Conservative 43; Mismatches 75; Indels 33; Gaps 9;

Qy 1 MCLNVKTLKGTNFEIEASPDASVADVKRIITETGQSTYRADQOQLYQKILKDETTLE 60  
Db 1 MKIFVKTLSGSNFEIEVKPADKVDKTALETYVG-AEYPAKQMLIHQGVKLDFTLE 59  
Qy 61 SNGVAENSFLVIMLSKAKASSSGASTAT---TAKAPATLAQPAAPVAPASVARTPTQA 116  
Db 60 ENNVVENSFVIMLSKTKASPGASTASAPSATQQTVAPOVS-APTASV----- 111  
Qy 117 PVATAETAPSPVQQAAPAAATVATDDADYVQAASNLVFGNNLEQTIQIILDMGGTWE 176  
Db 112 PVPTSGTA-----TAAAPA--TAASVQTVYVQAASNLVAGTTLESTVQIILDMGGSWD 164  
Qy 177 RDTVVRALRAAYNNPERAIDYLSGIPENVEAQPVAPAPAAAGQOTNOQASPAQPAVALP 236  
Db 165 RDTVVRALRAAFNNPERAVEYLSGIPQAQAEIPPAQAQATGEOANPLAQPOQAAA--- 221  
Qy 237 VQPSASAGNANPLNLPQGVPSGSGNPGVPGAGSGALDRLQLPQFOLLQVQANP 296  
Db 222 --PAATGGFNANPLNLPQGMFAADA-----GAGAGNLDFLNSQFQLRAMVQANP 273  
Qy 297 QILQPMLOELGKQNPQILRIQENQAEFLRLVNESPEGGFNGNIGLQAAAVPQTLVTVP 356  
Db 274 QILQPMLOELGKQNPQLVRLIQEADFLRLINEPVEGEB--NVMEQLEAMPQAVTVP 331  
Qy 357 EEREAIORLEGMGNRELVLVEVFACNCKDELTANYLLDHGHEPDDQ 403  
Db 332 EEREAIERLEGMGFDRAWVLEVFACNCKBELAANYLLDHMHFEFDQ 378

RESULT 3

AAG45204

ID AAG45204 standard; Protein; 378 AA.

XX

AC AAG45204;

XX 18-OCT-2000 (first entry)  
DT Arabidopsis thaliana protein fragment SEQ ID NO: 56723.  
XX  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 22-MAR-1999; 99US-0125788.  
PR 23-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137502.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.



```
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160743.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160988.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 58.9%; Score 1199.5; DB 21; Length 378;
Best Local Similarity 62.9%; Pred. NO. 5.8e-78;
Matches 256; Conservative 43; Mismatches 75; Indels 33; Gaps 9;

Oy 1 MKNVKTLLKGNFTEASPDASVADVKRIETTCQSTYRADQOMLIYQKILKDETTLE 60
Db 1 MKFVKTLLKGNFTEASPDASVADVKRIETTCQSTYRADQOMLIYQKILKDETTLE 59
Oy 61 SNGVAENSFLVIMLSKASGASTAT-----TAKAPATLAQPAAPVAPAPASVARTPTQA 116
Db 60 ENNVVENSFLVIMLSKASGASTASAPAPASPTQPTVATPQVS-APTASV----- 111
Oy 117 PVATAETAPPVQQAAPAAATVATDDADVYSQAASNLVFGNNLEQTIQQILDMGGGTWE 176
Db 112 FVPTSGTA-----TAAAPA--TAASVQTDVYQQAASNLVAGTILESTVQIILDMGGGSD 164
Oy 177 RDTVVRLAAVNNPERAIDLYSGIPENVEAQPVAPAPAGQQTNQQAASPAQPAVALP 236
Db 165 RDTVVRLAAVNNPERAVEYLYSGIPPAQAEIPPVQAQAPATGEQAANPLAQPOQAAA--- 221
Oy 237 VQSPASAGPNANPLNLPQCGVSGSGSNPGVVPGAGSGALDALRQLPQFQALLQLVQANP 296
```

```
Db 222 --PAAATGGPNANPLNLPQCGMPAADA-----GAGAGNLDFLNSQFQALRAMVQANP 273
Oy 297 QILOPMLQELGKQNPQILRIQENQAEFLRLVNESPRGGPGGNTLGGOLAAAVPOTLTVTTP 356
Db 274 QILOPMLQELGKQNPQILRIQEHQADFLRLINEPVEGER--NVMSQLEAAMPQAVTTP 331
Oy 357 EEREAIQRLEGMGFNRELVLVEFPACNKBELTANYLLDHGHEFDDQ 403
Db 332 EEREAIERLEGMGFDRAWVLVEFPACNKBELAAANYLLDHMHFEFDQ 378

RESULT 4
AAG19975
ID AAG19975 standard; Protein; 335 AA.
XX
AC AAG19975;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21982.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
```

PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 08-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 53.8%; Score 1095; DB 21; Length 335;  
Best Local Similarity 63.9%; Pred. No. 1.6e-70;  
Matches 232; Conservative 37; Mismatches 62; Indels 32; Gaps 8;

QY 45 MLIYQKILKDETTLESNGVAENSLVIMLSKAKASSSGASTAT-----TAKAPATLAQPA 100  
DB 1 MLIHQKVLKDETTLEENNVVENSFIVIMLSKTKASPSSGASTASAPASATQPTVATPQ 60  
QY 101 APVPAASAVARTPTQAPVATATETAPPSVQPOAATAVAATDDADVYSQAASNLVFGNNL 160  
DB 61 VS-APTASV-----PVPTSGTA-----TAAPA--TAASVQTDVYQAAASNLVAGTTL 105  
QY 161 ECTIQIILDMGGTWERDVTVRALRAAYNNPERAIDYLSGIPENVERAQPVARAPAAGQQ 220  
DB 106 ESTVQIILDMGGSDRDVTVRALRAAFNNPERAVEYLYSGIPAAEIPPPVAQAPATGEQ 165  
QY 221 TNQQAASPAQPAVALPVOPSPASAGPNANPLNPPQVPSGSGNPPGVVPGAGSGALDALR 280  
DB 166 AANPLAQPOQAAA-----PAAATGCPNANPLNPPQGMFAADA-----GAGAGNLDFLR 214  
QY 281 QLPOFQALLQVQANPQILOPMLQELGKONPQILRLLOENQAEFLRLVNESPEGGPGNI 340  
DB 215 NSQQFQALRAVQANPQILOPMLQELGKONPQILRLLOENQAEFLRLVNESPEGGPGNI -NV 272  
QY 341 LGQLAAAVPQTLVTPPEEREATQRLGNGFNRELVLVFFACNKDEELTANYLLDHGHEF 400  
DB 273 MEQLEAANPQAVTVTPPEEREATRLGNGFDRAMVLVFFACNKDEELTANYLLDHGHEF 332  
QY 401 DDQ 403  
DB 333 EDQ 335  
RESULT 5  
AAG45205  
ID AAG45205 standard; Protein; 335 AA.  
AC AAG45205;  
XX  
DT 18-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56724.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 23-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 29-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.



PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147152.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

PR	12-OCT-1999;	99US-0158369.	OS	Arabidopsis thaliana.
PR	13-OCT-1999;	99US-0159293.	XX	
PR	13-OCT-1999;	99US-0159294.	PN	EP1033405-A2.
PR	13-OCT-1999;	99US-0159295.	XX	
PR	14-OCT-1999;	99US-0159329.	PD	06-SEP-2000.
PP	14-OCT-1999;	99US-0159330.	XX	
PR	14-OCT-1999;	99US-0159331.	XX	25-FEB-2000; 2000EP-0301439.
PR	14-OCT-1999;	99US-0159637.	XX	25-FEB-1999;
PR	14-OCT-1999;	99US-0159638.	PR	99US-0121825;
PR	18-OCT-1999;	99US-0159584.	PR	99US-0123180.
PR	21-OCT-1999;	99US-0160741.	PR	99US-0123548.
PR	21-OCT-1999;	99US-0160767.	PR	99US-0125788.
PR	21-OCT-1999;	99US-0160768.	PR	99US-0126264.
PR	21-OCT-1999;	99US-0160770.	PR	99US-0126785.
PR	21-OCT-1999;	99US-0160814.	PR	99US-0127462.
PR	21-OCT-1999;	99US-0160815.	PR	99US-0128234.
PR	22-OCT-1999;	99US-0160980.	PR	99US-0128714.
PR	22-OCT-1999;	99US-0160981.	PR	99US-0129845.
PR	22-OCT-1999;	99US-0160989.	PR	99US-0130077.
PR	25-OCT-1999;	99US-0161404.	PR	99US-0130449.
PR	25-OCT-1999;	99US-0161405.	PR	99US-0130510.
PR	25-OCT-1999;	99US-0161406.	PR	99US-0130891.
PR	26-OCT-1999;	99US-0161359.	PR	99US-0131449.
PR	26-OCT-1999;	99US-0161360.	PR	99US-0132048.
PR	26-OCT-1999;	99US-0161361.	PR	99US-0132407.
PR	28-OCT-1999;	99US-0161920.	PR	99US-0132484.
PR	28-OCT-1999;	99US-0161992.	PR	99US-0132485.
PR	28-OCT-1999;	99US-0161993.	PR	99US-0132486.
PR	29-OCT-1999;	99US-0162142.	PR	99US-0132487.
Query Match 48.2%; Score 981; DB 21; Length 307;			PR	99US-0132863.
Best Local Similarity 62.7%; Pred. No. 2.le-62;			PR	99US-0134256.
Matches 210; Conservative 34; Mismatches 59; Indels 32; Gaps 8;			PR	99US-0134218.
			PR	99US-0134219.
QY	73	MLSKAKASSGASTAT-----TAKAPATLAQPAAPVAPAAASVARTPTQAPVATAETAPPSV 128	PR	99US-0134221.
Db	1	MLSKTASPSGASTASAPASATQPTVATPQVS-APTASV-----PVPTSGTA---- 48	PR	99US-0134370.
QY	129	QPOAAATVAATDDADVSOASNLVFGNNLEQTTQQLDMGGTWERDVTVVRALRAAY 188	PR	99US-0134378.
Db	49	-TAAAPF--TAAASQVDVYQQAASNLVAGTTLESTVQQLDMGGSWDRDVTVVRALRAAF 105	PR	99US-0134941.
QY	189	NNPERAIDLYSGIPENVEAQPVARAPAGQQTNQQAASPAQPAVALPVQPSPASAGPNA 248	PR	99US-0135124.
Db	106	NNPERAVEVLYSGIPAEIPFVQAAPATGEQANPLAQPOQAAA-----PRAATGSPNA 160	PR	99US-0135124.
QY	249	NPLNFPQGVPSGGSPGVVPGAGSGALDALROLQFQALLQLVQANPQILQPMQLQELGK 308	PR	99US-0135353.
Db	161	NPLNFPQGNPAADA-----GAGAGNLDFLRNSQQFQALRAMVQANPQILQPMQLQELGK 214	PR	99US-0135629.
QY	309	QNQQLRLIQENQAEFLRVNESPEGPGGNILGOLAAVPTLTVTPEERAIQBLEGM 368	PR	99US-0136021.
Db	215	QNQPLVRLQEQHQADEFRLINEPVEGEE--NVNMEQLEAAMPQAVTVPTEERAIERLEGM 272	PR	99US-0136392.
QY	369	GFNRELVLVEFFACNCKDEELTANYLLDHGHEFDQ 403	PR	99US-0136782.
Db	273	GFDRAMVLEVFFACNKEELAANYLLDHHEFDQ 307	PR	99US-0137222.
RESULT 7			PR	99US-0137528.
ID	AAG45206		PR	99US-0137502.
XX	AAG45206 standard; Protein; 307 AA.		PR	99US-0137724.
AC	AAG45206;		PR	99US-0138094.
XX			PR	99US-0138540.
XX	18-OCT-2000 (first entry)		PR	99US-0138847.
DT	Arabidopsis thaliana protein fragment SEQ ID NO: 56725.		PR	99US-0139119.
DE			PR	99US-0139452.
XX			PR	99US-0139453.
XX			PR	99US-0139454.
XX			PR	99US-0139455.
XX			PR	99US-0139456.
XX			PR	99US-0139457.
XX			PR	99US-0139458.
XX			PR	99US-0139459.
XX			PR	99US-0139460.
XX			PR	99US-0139461.
XX			PR	99US-0139462.
XX			PR	99US-0139463.
XX			PR	99US-0139750.
XX			PR	99US-0139763.
XX			PR	99US-0139817.
XX			PR	99US-0139899.
XX			PR	99US-0140353.
XX			PR	99US-0140354.
XX			PR	99US-0140695.
XX			PR	99US-0140823.
XX			PR	99US-0140991.

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

```

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 01-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159594.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 48.2%; Score 981; DB 21; Length 307;
Best Local Similarity 62.7%; Pred. No. 2.1e-62;
Matches 210; Conservative 34; Mismatches 59; Indels 32; Gaps 8;
Qy 73 MLSKAKASSGASTAT----TAKAPATLAQPAAPVAPAAASVARTPTQAPVATATAPPSV 128
Db 1 MLSKTKASPSGASTASAPASATQPTVATPQVS-APTASV-----PVPTSCTA---- 48
Qy 129 QPQAAPATVAATDDADVYSQAASNLVFGNNLEQTQQIILDMGGTWERDITVVRALRAAY 188
Db 49 -TAAAPA--TAASVQTDVYGOAASNLVAGTTLESTVQIILDMGGSGWDRTVVRALRAAF 105
Qy 189 NNPERAIDLYSGIPENVEAQPVARPAAGQQTNOQAASPAQPAVALPVOPSPASAGPNA 248
Db 106 NNPERAVEYLYSGIPQAQAEIPVQAQAPATGEQAANPLAQOQAAA-----PAAATGPNNA 160
Qy 249 NPLNLPQGVPSGSGNPGVVGAGSGALDALRQLPQFQALLQLVQANPQILQPMQLBLGK 308
Db 161 NPLNLPQGMFAADA-----GAGAGNLDPLRNSQQFQALRAMVQANPQILQPMQLBLGK 214
Qy 309 QNPQILRLIQENQAEFLRLVNESPEGPGGNIIGOLAAAVPQILTVTPPEREAIQRLEGM 368
Db 215 QNPQLVRLIOHQADFRLINEPEVEGEE--NVMEQLEAAMPQAVTVTPPEREAIQRLEGM 272
Qy 369 GFNRRLVLEVFACNKNDEULTANYLLDGHGFDDQ 403
Db 273 GFDRAMVLEVFACNKNDEELANYLLDGHMFEDQ 307

RESULT 8

```



AA017377  
ID AAG17377 standard; Protein; 371 AA.  
XX  
AC AAG17377;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 18372.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridization assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136352.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

PR	23-AUG-1999;	99US-0149902.	
PR	23-AUG-1999;	99US-0149930.	
PR	25-AUG-1999;	99US-0150566.	
PR	26-AUG-1999;	99US-0150884.	
PR	27-AUG-1999;	99US-0151065.	
PR	27-AUG-1999;	99US-0151066.	
PR	27-AUG-1999;	99US-0151080.	
PR	30-AUG-1999;	99US-0151303.	
PR	31-AUG-1999;	99US-0151438.	
PR	01-SEP-1999;	99US-0151930.	
PR	07-SEP-1999;	99US-0152363.	
PR	10-SEP-1999;	99US-0153070.	
PR	13-SEP-1999;	99US-0153758.	
PR	15-SEP-1999;	99US-0154018.	
PR	16-SEP-1999;	99US-0154039.	
PR	20-SEP-1999;	99US-0154779.	
PR	22-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	
PR	28-SEP-1999;	99US-0156458.	
PR	29-SEP-1999;	99US-0156596.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0158293.	
PR	13-OCT-1999;	99US-0158294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160780.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 46.4%; Score 945.5; DB 21; Length 371;			
Best Local Similarity 50.7%; Pred. No. 9.3e-60;			
Matches 208; Conservative 60; Mismatches 95; Indels 47; Gaps 11;			
QY	1	MKLNKTLKGTNFEIEASPDASVADVKRIETTGQSTYRADOQMLIYQGIKDKETTL 60	
DB	1	MKLTVTLKGSHFEIRVLPSTDMVKNKNIEDSQKDNYPCCQQLIHNGKVLKDKETSIV 60	
QY	61	SNGVAENGLVIMLSKAKASSGASTATTAKAPATLAQAPVAPAPASVARTPTQAPVAT 120	
DB	61	ENKVTGGLVVLMSKSK---SGGSAGQASVQTSSVQPV-----SATTSSTKPAAP-ST 111	
QY	121	ATAPSPVQPV---QAAPATVAATDADVYSQAASNLVFGNNLEQTIQQILDMGGCTWER 177	
DB	112	TOSSEVPASPIPAQCPAA-----QTDYTGQAASLVSGSLEQMVQVQIMEMGGGWDK 165	
QY	178	DTVTRALRAAYNNPERAIDLYSGIPENVE-AQPVARAPAGQQTNQQAASPAQPAVALP 236	

Db	166	ETVTRALRAAYNNPERAIDLYSGIPQTAEVAVVPEAQIAG-----SGAAP 212	
QY	237	VQSPASAGNANPLNLPQGVPSGGSGNPGVVPAGS---GALDALRQLPQFALLQLVQ 293	
Db	213	V--APASGGPNSSPLDLPQ-----ETVAAGSGDLGTLEFLRNDQFQQLRTMVH 261	
QY	294	ANPQILQPMQLQELGKQNPQILRLIQENQABFLRLVNESPEGGPG-GNILGQAAAAVPQT 352	
Db	262	SNPQILQPMQLQELGKQNPQILRLIQENQABFLQLVNEPYEGSDGEGDMPDQPEQEMPHAI 321	
QY	353	TVTPEBEATQRLGCGFNRELVLVVFACNKBELTANYLLDHGHEPDD 402	
Db	322	NVTPEBEATQRLGCGFNRELVLVVFACNKBELTANYLLDHGHEPDD 371	
RESULT 9			
ID	AAG44341	standard; Protein; 368 AA.	
XX	AC	AAG44341;	
XX	XX		
DT	18-OCT-2000	(first entry)	
XX	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 55531.	
XX	XX	Protein identification; signal transduction pathway; metabolic pathway;	
KW	KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	KW	termination sequence.	
XX	OS	Arabidopsis thaliana.	
XX	PN	EP1033405-A2.	
XX	PD	06-SEP-2000.	
XX	PF	25-FEB-2000; 2000EP-0301439.	
XX	PR	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	07-MAY-1999;	99US-0132487.	
PR	11-MAY-1999;	99US-0132863.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	

PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140895.  
PR 24-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 44.8%; Score 913; DB 21; Length 368;  
Best Local Similarity 47.5%; Pred. No. 2e-57;  
Matches 194; Conservative 64; Mismatches 104; Indels 46; Gaps 7;

Qy 1 MKNVTKLTGNTFEASPDASADVADVKRIETQSGSYRADQQMLIYQGIKLDKETTLE 60  
Db 1 MKNVTKLTGSHFELRVPTDIIMAVKNIEDSQKONYPGQQQLIHNGVKLDKETTLLV 60

Qy 61 SNGVAENSLVIMLSKAKASGASTATTAKAPATLAQPAAPVAPASVARTPTQAPVAT 120  
Db 61 ENKVTEGFLVWLSKSTASSGPSST-----QPTS-----TTTSTISST 101

Qy 121 AETAPPSVQPAAPATVA-----ATDDADVYSQAASHLVFGNNLEQTIQIILDMGGGTW 175  
Db 102 TLAAPSTTQSIAPVASNSTPVQEQPTAQSDTYGQAASLTVSGSSIEQMVQOIMEMGGGSM 161

Qy 176 ERDVTVRALRAAYNNPERAIDVLYSGIPENVEAQVARAPAGQQTNOQASPAQPAVAL 235  
Db 162 DKETVTRALRAAYNNPERAIDVLYSGIPETV---TIPATNLSGVSGSRELTAP----- 211

Qy 236 PVQSPASAGNANPLNLFPGVPSGGSNPGVVPAGSGALDALRQLPQFQALLQLVQAN 295  
Db 212 -----PPSGGNSPLDLFPQEAVSDA-----GGDLGTLEFLRGNDQFQQLRSMVNSN 260

Qy 296 PQILQPMLOELGKQNPQILRLIQENQASFLRLVNESPPGGPGG-NILGQLAAAVPQTUTV 354  
Db 261 PQILQPMLOELGKQNPQILRLIQENQASFLQLLNEPYEGSDGDVDIFQPDQEMPHSVNV 320

Qy 355 TPREEAATQRLGCMGFNRELVLVEFPACNDELTANYLLDHGHEFDD 402  
Db 321 TPEQESIERUEAMGFDRAIVIEAFLSCDRNEELAAANYLLEHSADFED 368

## RESULT 10

AA771459  
ID AA771459 standard; Protein; 368 AA.

AC AA771459;

DT 04-OCT-2000 (first entry)

DE Maize Rad23 protein #2.

KW Rad23; maize; ATCC No: PTA-531; recombinant expression cassette; wheat;  
transgenic plant; soybean; sunflower; sorghum; canola; modulator.

OS Zea mays.

PN WO200031268-A1.

PD 02-JUN-2000.

PF 12-OCT-1999; 99WO-US24129.

PR 23-NOV-1998; 98US-0109728.

PA (PION-) PIONEER HI-BRED INT INC.

PI Mahajan PB, Tagliani L;

DR WPI; 2000-400078/34.

DR N-PSDB; AAD01231.

PT Isolated nucleic acid encoding maize RAD23 protein is used to modulate  
the levels of polypeptides in plant or in assays for identifying  
compounds that bind to and/or increase/decrease enzymatic activity of  
catalytically active polypeptides -

PS Claim 11b; Page 78-79; 82pp; English.

CC The present sequence is the maize Rad23 protein #2. It is isolated from  
a Zea mays cell line, B73 callus tissue regenerated five days after  
transfer of the callus from medium containing auxin to a medium devoid

CC of exogenous auxin. The cDNA is deposited under the ATCC No: PTA-531.  
CC Maize Rad23 DNA sequence operably linked to a promoter, can be used to  
CC construct a recombinant expression cassette. This expression cassette  
CC can be used to generate a dicot or monocot transgenic plant e.g., maize,  
CC soybean, sunflower, sorghum, canola, wheat, etc. It can also be used to  
CC modulate the levels of Rad23 polypeptide expression in a plant or in  
CC assays to identify compounds, that bind to and/or modulate the enzymatic  
CC activity of catalytically active polypeptides.

SQ Sequence 368 AA;

Query Match 44.2%; Score 900.5; DB 21; Length 368;  
Best Local Similarity 49.0%; Pred. No. 1.6e-56;  
Matches 197; Conservative 53; Mismatches 117; Indels 35; Gaps 7;

Qy 1 MKNVTKLTGNTFEASPDASADVADVKRIETQSGSYRADQQMLIYQGIKLDKETTLE 60  
Db 1 MKNVTKLTGTHFIRVQPNDDTINAVKNTIEIQKDSYPWGQQLLHNGVKLDKETTLE 60

Qy 61 SNGVAENSLVIMLSKAKAS-SGASATATTAKAPATLAQPAAPVAPASVARTPTQAPVA 119  
Db 61 ENKVNEGFLVWLSKGTSGTSSQHSNTPATRAP-----PLEAPQAP-QPPVA 114

Qy 120 TATAPPSVQPAAPATVAATDDADVYSQAASHLVFGNNLEQTIQIILDMGGGTWERDT 179  
Db 115 PITTSQPEGLPAQAP-----NTHDNAASNLGSRNVDTIINOLMENGSGSWDKDK 164

Qy 180 VVRALRAAYNNPERAIDVLYSGIPENVEAQVARAPAGQQTNOQASPAQPAVALPQP 239  
Db 165 VQRALRAAYNNPERAIDVLYSGIPVTAB-----IAVIGGQANTTDRAPTGEA----- 213

Qy 240 SPASAGNANPLNLFPGVPSGGSNPGVVPAGSGALDALRQLPQFQALLQLVQANPQIL 299  
Db 214 -GLSGIPTAPDLDFPQASNAGG-----GAGGGLDFLRNNFQFQAVREMVHTNPQIL 266

Qy 300 QPMLOELGKQNPQILRLIQENQASFLRLVNESPPGGPGG-NILGQLAAAVPQTUTVPEER 359  
Db 267 QPMVELSKQNPQILRLIENHDFLQLLNEPFGEGEDFLDQPEEDEMPHAISVTPEEQ 326

Qy 360 EAIQRLGCMGFNRELVLVEFPACNDELTANYLLDHGHEFD 401  
Db 327 EAIQRLSGMGFDRAIVIEAFLACDRNEELAAANYLLEHAGEED 368

## RESULT 11

AA771378

ID AA771378 standard; Protein; 348 AA.

AC AA771378;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 18373.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.

```
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 43.3%; Score 881.5; DB 21; Length 348;
Best Local Similarity 50.6%; Pred. No. 3.3e-55;
Matches 195; Conservative 56; Mismatches 87; Indels 47; Gaps 11;

QY 26 VKRIITETGQSTYRADQOQMLYQGIKIDETTLSENGVAENSLVIMLSKAKASSGAS 85
DB 3 VKKNIEDSQKDNPCGQQLIHNGKVLKDETSLENKVTBERGFLVWMLSKS---SGGS 59
QY 86 TATTAKAPATLAQAPAPVAPASVARTFQAPVATETAPPSVQP---QAAPAAVTAATD 142
DB 60 AQGASVQTSVSPQPV-----SATTSTKPAAP-STTQSSPVSPASPIPAQEQPAA----- 107
QY 143 DADVTSQAASNLVFGNNLEPOTIQIILDMGGGTWERDVTVALRAAYNNPERAIDVLYSGI 202
DB 108 QTDYTGQAASTLVSSLSQWQCIEMEGGSGNDKETVTVALRAAYNNPERAVDLYLYSGI 167
QY 203 PENVE-AQVPAPAPAGQOTNOQAAAPQAPVALVPQSPASAGPNANPLNLPQGVFSG 261
DB 168 PQTAEVAVVPEAQIAG-----SCAAPV--APASGGPNSSPLDLFPQ----- 207
QY 262 GSNPGVVPAGS---GALDALQLQFQALLQLVQANPOILOPMLQELGKQNPQILRLIQ 318
DB 208 ----ETVAAAGSGDLGTLLEFLRNNDFQQLRTWVHNSNPQILQPMQLQELGKQNPQLRLIQ 263
QY 319 ENQASFELRLVNESPEGGPG-GNILGQAAAAPQTLVTPPEEATQRLGEGMGNRELVL 377
DB 264 ENQAEFLQLVNFPEYSGSDGDMFQPBQEMPHAINVTPEAQEQAIQRLGEMGFDRLVIE 323
QY 378 VFFACNKBELTANTYLLDHGFDD 402
DB 324 AFLACDRNELAANYLLENSGDFED 348
```

```
RESULT 12
AAG44342
ID AAG44342 standard; Protein; 345 AA.
XX AC AAG44342;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 55532.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
```

```
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
```







```
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 37.1%; Score 754.5; DB 21; Length 299;
Best Local Similarity 50.3%; Pred. No. 3.5e-46;
Matches 170; Conservative 48; Mismatches 73; Indels 47; Gaps 11;

Qy 73 MLSKASSSGASTATTAKAPATIAQAPVAPASVARTQAPVATAETAPSVOP-- 130
Db 1 MLSKSK---SGSGAGASVQTSVSQPV-----SATTSSSTKPAAP-STTQSSVPVAPSP 51
Qy 131 -QAAPAAATVATDADVYSQAASNLVFGNNLEQTIQOILDMGGGTWERDVTVRALRAAYN 189
Db 52 AQEQPAA-----QTDYTGQAATLVSGSSLEQVQIMENGSGGSDKKEITVRALRAAYN 105

Qy 190 NPERAIDLYSGIPENVE-AQPVARAPAAAGQQTNOQAASPAQPAVALPVQSPSPASGPN 248
Db 106 NPERAVDLYSGIPQTAEVAVPVPEAQTAG-----SGAAPV--APASGGPN 150
Qy 249 NPLNLFPPQGVSGGSGGPNVPGAGS---GALDALRQLPQFQALLQLVQANPQILQPMLOE 305
Db 151 SPLDLFPQ-----ETVAAAGSGDLGTLEFLRNDDQFQLRTVMHNSNPQILQPMLOE 201
Qy 306 LKQNPQILRLIOENQAEFLRLVNESPEGGPG-GNIGLQALAAVPTLTTTPEERAIQR 364
Db 202 LKQNPQLRLIOENQAEFLRLVNEPYEGSGDEGDFQEQEMPHAINVTFAEQEAIQR 261
Qy 365 LEGMGFNRELVLVEFFACNCKDELTANYLLDHGHEFDD 402
Db 262 LEAMGFDRALVIEAFLACDRNEELANYLLENSGDFED 299

RESULT 14
AAG36529
ID AAG36529 standard; Protein; 257 AA.
XX
AC AAG36529;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 44779.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
```

PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 17-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139760.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.

[illegible]

```
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
```

```
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 35.6%; Score 725; DB 21; Length 296;
Best Local Similarity 46.4%; Pred. No. 4.6e-44;
Matches 156; Conservative 53; Mismatches 81; Indels 46; Gaps 7;

Qy 73 MLSKAKASSGASTATTAKAPATLAQPAAPVAPASVARTPTQAPVATAETAPSPVQQA 132
Db 1 MLSKSKTASSAGPSST-----QPTS-----TTSTISSTLLAAPSTQSI 41
Qy 133 APAATVA-----ATDDADVYSQAASNLVFGNNLEQTIQQIILDMGGGTWERDVTVRALRAA 187
Db 42 VPASNSTPVQEQPTAQSDTYGQAASSTLVSGSSIEQMVOQIEMGGGSDKETVTRALRAA 101
Qy 188 YNNPERAIDYLSGIPENVEAQPVARAPAAQOQTNQQAASPAQPAVALPVQPSPASGPN 247
Db 102 YNNPERAVDLYSGIPETV---TIPATNLGSGVSGRELTAP-----PPSGGPN 146
Qy 248 ANPLNLPQGVPSGSGNPGVPGAGSGALDALROLPOFOALLQLVQANPQILOPMLQELG 307
Db 147 SSPLDLFPQEAVDAA-----GGDLGTLFIRGNDQFQQLRSMVNSNPQILOPMLQELG 200
Qy 308 KONPQILRLIQENQAELRLVNESPEGGPGG-NILGOLAAVPOTLTVPPEERAIORLE 366
Db 201 KONPQILRLIQENQAELRLVNESPEGGPGG-NILGOLAAVPOTLTVPPEERAIORLE 260
Qy 367 GMGFNRELVLVFFACNKBDELTANYLLDHGHEFDD 402
Db 261 AMGFDRIVIEAFLSCDRNEELAAANYLLEHSADFED 296
```

Search completed: December 17, 2003, 06:18:38  
Job time : 40.771 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:20:11 ; Search time 138.842 Seconds  
(without alignments)  
544.771 Million cell updates/sec

Title: US-09-805-550-2

Perfect score: 2036  
Sequence: 1 MKNVTKLGTNFIEASPD.....BELTANYLLDHGHEFDQQQ 405

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2036	100.0	368	9	US-09-805-550-2
2	900.5	44.2	405	9	US-09-805-550-2
3	573	28.1	363	12	US-10-116-275-170
4	572.5	28.1	379	15	US-10-102-806-549
5	198	9.7	595	12	US-10-293-000-2
6	197	9.7	589	12	US-10-293-000-4
7	166.5	8.2	624	15	US-10-146-473-47
8	151	7.4	79	11	US-09-918-036-4
9	149	7.3	1023	11	US-09-893-519A-14
10	146	7.2	681	9	US-09-815-242-11830
11	144	7.1	727	15	US-10-234-432-30
12	144	7.1	727	15	US-10-234-432-57
13	141.5	6.9	638	15	US-10-234-432-59
14	141.5	6.9	2971	15	US-10-146-473-50
15	140	6.9	2441	14	US-10-109-886-8

16	138.5	6.8	522	15	US-10-296-770-2	Sequence 2, Appli
17	138	6.8	81	11	US-09-918-036-5	Sequence 5, Appli
18	137	6.7	578	12	US-10-032-585-7793	Sequence 7793, Ap
19	137	6.7	810	15	US-10-156-761-10081	Sequence 10081, A
20	136	6.7	655	14	US-10-001-632A-2	Sequence 2, Appli
21	136	6.7	655	15	US-10-054-683-29	Sequence 29, Appli
22	135	6.6	352	11	US-09-820-843A-23	Sequence 23, Appli
23	134.5	6.6	699	12	US-10-200-562-143	Sequence 143, App
24	134.5	6.6	699	12	US-10-237-551-143	Sequence 143, App
25	134.5	6.6	699	12	US-10-237-551-254	Sequence 254, App
26	134.5	6.6	699	15	US-10-121-988-143	Sequence 143, App
27	134.5	6.6	1276	15	US-10-156-761-10509	Sequence 10509, A
28	134	6.6	1004	10	US-09-738-626-5676	Sequence 5676, Ap
29	133.5	6.6	358	12	US-10-024-298A-137	Sequence 137, App
30	133.5	6.6	358	12	US-10-042-211A-137	Sequence 137, App
31	133.5	6.6	358	15	US-10-197-666A-32	Sequence 32, Appli
32	133.5	6.6	396	12	US-10-024-298A-139	Sequence 139, App
33	133.5	6.6	396	12	US-10-042-211A-139	Sequence 139, App
34	133.5	6.6	396	15	US-10-197-666A-34	Sequence 34, Appli
35	133.5	6.6	718	12	US-10-029-386-34049	Sequence 34049, A
36	133.5	6.6	2442	14	US-10-109-886-10	Sequence 10, Appli
37	132.5	6.5	309	11	US-09-820-843A-24	Sequence 24, Appli
38	132	6.5	1259	15	US-10-260-715-8	Sequence 8, Appli
39	131.5	6.5	1744	14	US-10-108-605-25	Sequence 25, Appli
40	131	6.4	1247	15	US-10-128-714-3473	Sequence 3473, Ap
41	131	6.4	1263	15	US-10-128-714-8473	Sequence 8473, Ap
42	130	6.4	2382	15	US-10-196-935A-2	Sequence 2, Appli
43	129	6.3	63	9	US-09-925-299-1544	Sequence 1544, Ap
44	129	6.3	63	11	US-09-925-299-1544	Sequence 1544, Ap
45	129	6.3	504	11	US-09-938-864-410	Sequence 410, App

#### ALIGNMENTS

#### RESULT 1

US-09-805-550-2  
; Sequence 2, Application US/09805550  
; Patent No. US20020026045A1  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; APPLICANT: Tagliani, Laura  
; TITLE OF INVENTION: Rad23 Genes and Uses Thereof  
; FILE REFERENCE: 0964D  
; CURRENT APPLICATION NUMBER: US/09/805,550  
; CURRENT FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 09/413,574  
; PRIOR FILING DATE: 1999-10-06  
; PRIOR APPLICATION NUMBER: 60/109,728  
; PRIOR FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 405  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-805-550-2

Query Match	100.0%;	Score	2036;	DB	9;	Length	405;
Best Local Similarity	100.0%;	Pred. No.	2.3e-146;				
Matches	405;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MKNVTKLGTNFIEASPDASVADVKRIITETGOSTYRADQOMLIYQKILKDETTLE	60				
Db	1	MKNVTKLGTNFIEASPDASVADVKRIITETGOSTYRADQOMLIYQKILKDETTLE	60				
Qy	61	SNGVAENFLVIMLSKAKASSGASTATTAKAPATLAQAPAPVAPAPASVARTPTQAPVAT	120				
Db	61	SNGVAENFLVIMLSKAKASSGASTATTAKAPATLAQAPAPVAPAPASVARTPTQAPVAT	120				
Qy	121	AETAPPSVQQAAPAAVATDDADVYSQAASNLVFGNNLEQTIQQIILDMGGGTTWERDVT	180				
Db	121	AETAPPSVQQAAPAAVATDDADVYSQAASNLVFGNNLEQTIQQIILDMGGGTTWERDVT	180				

Qy 181 VRALRAAYNNPERAIDLYSGIPENVEAQVAPAPAGQQTNOQAASPAQVALPVQPS 240  
Db 181 VRALRAAYNNPERAIDLYSGIPENVEAQVAPAPAGQQTNOQAASPAQVALPVQPS 240  
Qy 241 PASAGPNANPLNFPQGVPSGSGNPGVPGAGSGALDRLQLPQFQALLQVQANPQILQ 300  
Db 241 PASAGPNANPLNFPQGVPSGSGNPGVPGAGSGALDRLQLPQFQALLQVQANPQILQ 300  
Qy 301 PMLQELGKQNPQILRLIQENQAEFLRLVNSPEGGPGGNILGOLAAAVPOTLTVTPERE 360  
Db 301 PMLQELGKQNPQILRLIQENQAEFLRLVNSPEGGPGGNILGOLAAAVPOTLTVTPERE 360  
Qy 361 AIQRLGGMGNRELVLVFFACNKKDELTANYLLDHGHEFDQOQ 405  
Db 361 AIQRLGGMGNRELVLVFFACNKKDELTANYLLDHGHEFDQOQ 405

RESULT 2  
US-09-805-550-4  
; Sequence 4, Application US/09805550  
; Patent No. US2002026045A1  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; TITLE OF INVENTION: Rad23 Genes and Uses Thereof  
; FILE REFERENCE: 0964D  
; CURRENT APPLICATION NUMBER: US/09/805,550  
; CURRENT FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 09/413,574  
; PRIOR FILING DATE: 1999-10-06  
; PRIOR APPLICATION NUMBER: 60/109,728  
; PRIOR FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-805-550-4

Query Match 44.2%; Score 900.5; DB 9; Length 368;  
Best Local Similarity 49.0%; Pred. No. 2.8e-60;  
Matches 197; Conservative 53; Mismatches 117; Indels 35; Gaps 7;

Qy 1 MKNVTKLKTGNFEIEASPDASVADVKRIETTTQGSTYRADQOQMLIYQKILKDETTLE 60  
Db 1 MKLTVTKLKTGTHFEIRVQPNDTIMAVKKNIIEIQKDSYPMGQQLLIFNGKVLKDESTLE 60

Qy 61 SNGVAENSLVIMLSKAKAS--SSGASTATTAKAPATLAQAPAPVAPAAASVARTPTQAPVA 119  
Db 61 ENKVNEDGFLVVMLSKGTSGTSSQHSNTPTATQAP-----PLEAQOAP-QPPVA 114

Qy 120 TAETAPSVQQAAPAAATDADVYSQAASNLVFGNNLEQITIQIILDMGGGTWERDT 179  
Db 115 PITTSQPEGLPAQAP-----NTHDNAASLLSGNRVDTIINLMEMGGSGWKDK 164

Qy 180 VVRLRAAYNNPERAIDLYSGIPENVEAQVAPAPAGQQTNOQAASPAQVALPVQPS 239  
Db 165 VQRLRAAYNNPERAVEYLYSGIPVTAE-----IAVPICGGQANTTDRAPTGEA----- 213

Qy 240 SPASAGPNANPLNFPQGVPSGSGNPGVPGAGSGALDRLQLPQFQALLQVQANPQIL 299  
Db 214 -GLSGIPNTAPDLDFPQASNAGG-----GAGGGPLDFLRNNPFOQAVREMTNPNQIL 266

Qy 300 QPMLQELGKQNPQILRLIQENQAEFLRLVNSPEGGPGGNILGOLAAAVPOTLTVTPERE 359  
Db 267 QPMLVELSKQNPQILRLIEENHDFLQLLNPFEGGEGDFLDQPEDEMHAISVTPEEQ 326

Qy 360 EAIQRLGGMGNRELVLVFFACNKKDELTANYLLDHGHEFD 401  
Db 327 EAIQRLGSMGDFRARVIEAFLACDRNEELAAANYLLLEHAGEED 368

RESULT 3  
US-10-116-275-170  
; Sequence 170, Application US/10116275  
; Publication No. US20030211476A1  
; GENERAL INFORMATION:  
; APPLICANT: Eian Pharmaceutical Technology  
; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Brayden, David  
; APPLICANT: Byrne, Daragh  
; APPLICANT: Lambkin, Inelda  
; APPLICANT: Higgins, Lisa  
; TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and Methods an  
; FILE REFERENCE: E1067/20087  
; CURRENT APPLICATION NUMBER: US/10/116,275  
; CURRENT FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 349  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 170  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-116-275-170

Query Match 28.1%; Score 573; DB 12; Length 363;  
Best Local Similarity 33.3%; Pred. No. 2e-35;  
Matches 139; Conservative 79; Mismatches 124; Indels 76; Gaps 13;

Qy 3 LNVTKLKTGNFEIEASPDASVADVKRIETTTQGSTYRADQOQMLIYQKILKDETTLESN 62  
Db 5 ITLTKLQQOTFKIRMEPDETIVKLEBAEKGRDAPPVAGQKLIYAGKILSDDDVPIRDY 64

Qy 63 GVAENSLVIMLSKAKAS--SSGASTATTAKAPATLAQAPAA---PVAPAAASVARTPTQAPVA 119  
Db 65 RIDERKFNVMVMTKTA---GQTSAPPEASPTAPESSTSPPTSGMSHPP---PAA 118

Qy 120 TAETAP-----PSVQQAAPAAATDADVYSQAASNLVFGNNLEQITIQIILDMGGGT 174  
Db 119 REDKSPSEASAPTSPEV-SGSVPSSGSGREDDAASLTIVTSEYETMLTMSMG--- 174

Qy 175 WERTTVVRLRAAYNNPERAIDLYSGIPENVEAQVAPAPAGQ-QTNQQAASPAQPAV 233  
Db 175 YERERVVAALRASYNPHRAVEYLLTGIPGSP-----PEHGSVQESQVSEQPATEA- 226

Qy 234 ALPVQSPASAGPNANPLNFPQGVPSGSGNPGVPGAGSGALDRLQLPQFQALLQVQ 293  
Db 227 -----AGENPLEFLRDPQFQONMRQVIQ 249

Qy 294 ANPQILQPMLOELGKQNPQILRLIQENQAEFLRLVNSPEGGPGGNI-----LGQLAAA 347  
Db 250 QNPALLPALLQQLGOENPQLLQOISRHOQFIQMLNEPP--GELADISDVEGEVGAIGEE 307

Qy 348 VPQT--LTVTPEREAIQRLGGMGNRELVLVFFACNKKDELTANYLLDHGHEFDQO 403  
Db 308 APQWNYIQTVPQEKAIERLKGFPPESLVIOAYFACEKNEKNAANPLL--SQNFDDDE 363

RESULT 4  
US-10-102-806-549  
; Sequence 549, Application US/10102806  
; Publication No. US2003005421A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA103P1C1  
; CURRENT APPLICATION NUMBER: US/10/102,806  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/925,298  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05881  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270





Db 153 GGLAGLSGLNTNFSELSQ-----MORQLLSNPEMMYQIMENPFVQSMLSNPDLMRQ 207  
QY 154 LVFGN-NLEQIQIILDMGGTWERDTTVRALRAAYN-----NPERADLYSGIP 203  
Db 208 LIMANFQMQLIQRNPEISHMLNPNPDMQTLERARNPAMQEMMRNQRLSNLES-IP 266  
QY 204 ENVEA-----QPVARA-----PAAGQOTNQQAASPAQPAVALFVOPSPASAGP 246  
Db 267 GGYNALRRMYTDIQEPMLSAAQFGNPPASIVSNTSSGEGSQPSRTENRDLPLNPAP 326  
QY 247 NANPLNLPFGVPS--GGSN-----PGVPGAGSGALDALRQLPQFOALLQ 290  
Db 327 QTSQSSASSTASTVGGTTGSGTASGTSGSTTAPNLVPGVGASMENT-----PGMSLIQ 382  
QY 291 LVQANPQILQ-----PMLQELGKQ----- 309  
Db 383 QITENPQLMQLMSAPYMRSMQSLSONPDLAAQMLNPNPLFAGNPQLQEQMRQQLPTFL 442  
QY 310 -----NPOILRIQENQAEFLRLVNESP-----EGGPGG--- 338  
Db 443 QMQNPDTLSAMSNPRAMQALLIQOQLQTLATEAPGLIPGFTPGGLGALGSTGSSGTNG 502  
QY 339 -----NILGOLAAAVPOTLTVTPEEREAIQRLGEMGF-NR 372  
Db 503 SNATPSENTSPTAGTTEPGHQFIQMLQALAGVNPQLQNPVRFQOQLEQLSANGFLNR 562  
QY 373 ELVLEVPFACNKBELTANVLL 394  
Db 563 EANLQALATGGBDINAARILL 584

RESULT 7  
US-10-146-473-47  
; Sequence 47, Application US/10146473  
; Publication No. US20030108888A1  
; GENERAL INFORMATION:  
; APPLICANT: Scanlan, Matthew  
; APPLICANT: Gout, Ivan  
; APPLICANT: Stockert, Elisabeth  
; APPLICANT: Gure, Ali  
; APPLICANT: Chen, Yao-Tseeng  
; APPLICANT: Old, Lloyd  
; TITLE OF INVENTION: Breast Cancer Antigens  
; FILE REFERENCE: L00461/70130(JRV)  
; CURRENT APPLICATION NUMBER: US/10/146,473  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/291,150  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn version 3.0.0  
; SEQ ID NO 47  
; LENGTH: 624  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-146-473-47

Query Match 8.2%; Score 166.5; DB 15; Length 624;  
Best Local Similarity 22.2%; Pred. No. 0.00028;  
Matches 100; Conservative 64; Mismatches 178; Indels 109; Gaps 20;  
QY 1 MKNLVKTLKGTNFEIASPDASVADVVKRIETTTQGSTYRADQOOLTYQKILKDETTLE 60  
Db 33 IKVTVPKKEB-EFAPVENSQVQFKEAIKRFKSGT---DQLVLIFAGKILKQDQTLI 88  
QY 61 SNGVAENFLVIMLSKAKASSSGASTATTAKAPATLQAAPVAPAASVARTP--TQAPV 118  
Db 89 QHGI-HDGLTVHLVIKSNRPQSGT-----QPSNAAGTWTTSASTRSNSTPI 136  
QY 119 AT-----AETAPPVQVQ-----AAPAATVAATDDADVYSQAA 151  
Db 137 STNSNPFGLSGLAGLSLGLSSTNFSLSQMQOQLMASPEMMIQIMENPFVQSMLS 196

QY 152 S-----NLVFGN-NLEQTIQ-----QILDMGGTWERDTTVRALRAAYN 189  
Db 197 NPDLMRQLIMANPQMQLIQRNPEISHMLNPNPDMQTLERARNPAMQEMMRNQRLALS 256  
QY 190 NPE-----RAIDYLYSGIPE---NVEAQPVARAPAAQOQTNQQAASPAQPA-----V 233  
Db 257 NLESTPGGVNLRMYTDIQEPMLSAAQFGNPPASIVSNTSSGEGSQPSRTENRDLPL 316  
QY 234 ALPVOPSPASAPNANPLNLPQGVPSGSGSNPVGFG--AGSGALDALRQLPQFOALLQ 291  
Db 317 PNPWAPPATQS-SATTSTTTSTGSGSGSNATGNTVAAANYVASIFSTPGMSLLQ 375  
QY 292 VQANPQILQPM-----QELGKQPQILRIQENQAEFLRLVNESPEGGPGGNILG 342  
Db 376 ITENPQLIQNMLSAFYMRSMMQSL--SONPDLAAQMLNPLFTANPQLQEQMRP----- 428  
QY 343 QLAHAV-----PQTLTV--TPEEREAIQRL 366  
Db 429 QLPAPFLQMQNPDTLSAMSNPRAMQALMIQ 459

RESULT 8  
US-09-918-036-4  
; Sequence 4, Application US/09918036  
; Publication No. US20030092159A1  
; GENERAL INFORMATION:  
; APPLICANT: MADURA, Kitan  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE RAPID PURIFICATION OF PROTEAS  
; FILE REFERENCE: 266/165  
; CURRENT APPLICATION NUMBER: US/09/918,036  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: 60/050,171  
; PRIOR FILING DATE: 1997-06-19  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 79  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-918-036-4

Query Match 7.4%; Score 151; DB 11; Length 79;  
Best Local Similarity 38.0%; Pred. No. 0.00026;  
Matches 30; Conservative 21; Mismatches 28; Indels 0; Gaps 0;  
QY 1 MKNLVKTLKGTNFEIASPDASVADVVKRIETTTQGSTYRADQOOLTYQKILKDETTLE 60  
Db 1 MOVTLKTLQOQTFKIDIDPEETVKALKEIESEKGDAPPVAGOKLIYAGKILNDDTALK 60  
QY 61 SNGVAENFLVIMLSKAKA 79  
Db 61 EYKIDKPNFVVVWVTKPKA 79

RESULT 9  
US-09-893-519A-14  
; Sequence 14, Application US/09893519A  
; Publication No. US20030027243A1  
; GENERAL INFORMATION:  
; APPLICANT: ANADYS PHARMACEUTICALS, INC.  
; APPLICANT: THOMPSON, Craig  
; APPLICANT: MOORE, Jeffrey  
; APPLICANT: BUURMAN, Ed T.  
; APPLICANT: BRADLEY, John  
; APPLICANT: DESILVA, Thamara  
; APPLICANT: HARRIS, Sandra  
; APPLICANT: KOMARNITSKY, Svetlana  
; APPLICANT: MENDELLO, Marc  
; APPLICANT: MOORE, Daniel  
; APPLICANT: MCCOY, Melissa  
; APPLICANT: SANDERSON, Karen  
; APPLICANT: HAQ, Tariq

; APPLICANT: ZHU, Shuhao  
; APPLICANT: LONG, Fan  
; APPLICANT: DAVIDOV, Eugene  
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE  
; FILE REFERENCE: 0342/1G548-US2  
; CURRENT APPLICATION NUMBER: US/09/893,519A  
; PRIOR FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: US 60/215,164  
; PRIOR FILING DATE: 2000-06-29  
; PRIOR APPLICATION NUMBER: US 60/224,457  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 146  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; TYPE: PRT  
; LENGTH: 1023  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Corresponds to SEQ ID NO: 87  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: Human Genbank/CAA72189  
; DATABASE ENTRY DATE: 1997-06-25  
; RELEVANT RESIDUES: (1)..(1023)  
US-09-893-519A-14

Query Match 7.3%; Score 149; DB 11; Length 1023;

Best Local Similarity 26.1%; Pred. No. 0.012; Indels 84; Gaps 16;  
Matches 86; Conservative 31; Mismatches 128;

Qy 58 TLENGVAENSLVIMLSKAKSSGASTATT-----AKAPATLAQPAAPVAPA-ASVART 112  
Db 214 SLVNGPA-----ALLPLPKPAAPGTVIQTTPFVGAAAPAPAPAPSPAPAPAPAAA 267  
Qy 113 PTQAPVATETAPSPVQQAAPATVAATDDADVYSQAASNLVFGNNLEQTIQIILDMGG 172  
Db 268 PPPPPAPATLARPPGHPAGPPTAPAVPPPA-----AAQN-----G 304  
Qy 173 GTWERDVTVRALRAAYNNPERAIDLYSGIPENVEAQP-----VARAPAGOOTNQ----- 223  
Db 305 GS-----AGNAPAPAPA-----AGGAGVSGQPGPGGAAAAPAGVKAESPKEV 349  
Qy 224 QAASPAQPAVALPVQSPASA-----GPNANPLNLPQGV-----PSGSGNPGVVPFGAGSALDA 278  
Db 350 QAAPPAATLAASGASTAASVIGTQWQALPSPAAPPVPPAPGTPTGLPKGAAGAVTQS 409  
Qy 279 LRQLPQ-----FOALLQLVQANPQILQPMLOELKQNPQI-----LRLIQENQAEFLRLVNE 330  
Db 410 LSRTPTATTSGIRATLTPTVLAPLPQPPQNPNTNQNFPQLPPGMVLRSENGOLLMTIPQ 469  
Qy 331 SPEGGPGGNILGOLAA-----AVPQTLTVTP 356  
Db 470 A-----LAQMOQAHAQPPQT-TMAP 488

## RESULT 10

US-09-815-242-11830  
; Sequence 11830, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11830  
; LENGTH: 681  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-11830

Query Match 7.2%; Score 146; DB 9; Length 681;

Best Local Similarity 23.7%; Pred. No. 0.011; Indels 96; Gaps 18;  
Matches 98; Conservative 56; Mismatches 164;

Qy 16 EASPDPA---SVADVKKRIETTGQSTYRADOQMLIVQGIKDKDTTLESNGVAENSLVI 72  
Db 296 QALPEALDNGQGRERVIALAQAALPA---EDVQFYQWGLIGRRDLPLAPD---PRSGFEMV 351  
Qy 73 ML-----SKAKASSGASTATTAKAPATLAQPAAPVAPAASVARTPTQAPVA 119  
Db 352 LLRLMLAFRPADAGVPRPTPLKDLGISKATTDPANSPVAGASP-APVATVAPVPAAPV 410  
Qy 120 TATAPSPVQQAAPATVAATDDADVYSQAASNLVFGNNLEQTIQIILDMGGTWER-- 177  
Db 411 EAPAAPPAAP-PSAPPAAPAAVEARVTEAVVVEEPAAA-----AEVVDL---PWSEPA 454  
Qy 178 -----DTVVRALRAAYNNPERAIDLYSGIPENVEAQPVARAPAGOOTNQAA 226  
Db 455 PSLAAPEPEPEPEPEPLAVEAPSPVPAVAVAVETV---LEALPAALPVAPDEQDEQ 511  
Qy 227 SP-----AQPAVALPVQSPSA-----SAGPNANPLNLPQ-----GVPS 260  
Db 512 PPPADDYVEVDMDTLAYLDATPEPDVVVVEEPLPAAPKATGLAAEWLELPRGLGLGITA 571  
Qy 261 G-GSNPGVV-----PGAGSALDALQOLPOFQALLQLVQANPQILQPMLOELG 307  
Db 572 SIGANCTLVAAADDDHWHHLDPGQ-SALFNATQORRLNDALNQHGLRTLK-LEVTLQKPE 629  
Qy 308 KQNPQILRLIQENQAEFLRLVNESPEGGPGGNILGOLAAAVPOTLTVTTPERE 361  
Db 630 QETP---AQAAARRRAERQRAAEASIDADPLVRQLREOPAAVVRDGTIEPLEAKA 681

## RESULT 11

US-10-234-432-30  
; Sequence 30, Application US/10234432  
; Publication No. US20030091598A1  
; GENERAL INFORMATION:  
; APPLICANT: Homer, Mary J.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Persing, David H.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
; TITLE OF INVENTION: TREATMENT OF BABESIA INFECTION  
; FILE REFERENCE: 210121.560  
; CURRENT APPLICATION NUMBER: US/10/234,432  
; CURRENT FILING DATE: 2002-08-30  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: Corixa Invention Disclosure Database  
; SEQ ID NO 30

```
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Babesia sp. WA1
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 264
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-234-432-30

Query Match          7.1%; Score 144; DB 15; Length 727;
Best Local Similarity 24.4%; Pred. No. 0.018;
Matches 95; Conservative 45; Mismatches 169; Indels 80; Gaps 16;

Qy 17 ASDASVADVKRIETTGQSTYRADQOQMLIYQKILKDETTLESNGVAENSLFLVIMLSK 76
Db 72 ATPEASQGSANQ-----QQSQTGAGESQPVLSSTEMATVKEETVPETK--VENNVV-----120

Qy 77 AKASSSGASTATTAKAPA--TLAQ--PAAPVAPAASVARTPTQ--APVATAETAPPSVQPOA 132
Db 121 -----QEATVTPAQVPAVENVSQPTQTVAAPAAPQPAQVAPQATAGIQQAQPPVA 174

Qy 133 APAAT-----VAATDDADVISOAASN-----LVFGNNLEQTIQIILDMGGGTWER 177
Db 175 TETATAEQPVAAATTTVEQMPQAAASPAPILETPOVMTQTAPVEETQAPVVTESPAQOP 234

Qy 178 DTVVRALRAAYNNPERAIDLYSGIPENVEAQVAPARAAGQQTNQQAASPAQPAVALPV 237
Db 235 AQV-----AAPEQPAEVAPQATAGI--QQAQOPQVAVATETATAEQPVAAATTTVEQMPQAAAE 288

Qy 238 QPSPASAGPNA-----NPLNLFPGQVPSGSGSNPGVVPAGSGGALDALR 280
Db 289 SPAPISSETPOVMTQTAPVEETQAPVVTESPAQPAQVAAPEQPAEVAPQATAGIQ--344

Qy 281 QLPQFQALLQVQANPOILOPMLQELGKQ---NPQILRLIQENQAEFLRLVNESPEGGPG 337
Db 345 -----QAQOPQVAAEAQVQPPVQTAQTRPVAQPV--VVAEAQ-----VVQPPVKAQA 391

Qy 338 GNIILGQLAAAVPQTLTVPREAIQRL 366
Db 392 AQPVVKDQAQP--VASVAPQATAGIQQAQ 419

RESULT 12
US-10-234-432-57
; Sequence 57, Application US/10234432
; Publication No. US20030091598A1
; GENERAL INFORMATION:
; APPLICANT: Homer, Mary J.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Persing, David H.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF BABESIA INFECTION
; FILE REFERENCE: 210121.560
; CURRENT APPLICATION NUMBER: US/10/234,432
; CURRENT FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 57
; TYPE: PRT
; ORGANISM: Babesia sp. WA1
US-10-234-432-57

Query Match          7.1%; Score 144; DB 15; Length 727;
Best Local Similarity 24.4%; Pred. No. 0.018;
Matches 95; Conservative 45; Mismatches 169; Indels 80; Gaps 16;

Qy 17 ASDASVADVKRIETTGQSTYRADQOQMLIYQKILKDETTLESNGVAENSLFLVIMLSK 76
Db 72 ATPEASQGSANQ-----QQSQTGAGESQPVLSSTEMATVKEETVPETK--VENNVV-----120

Qy 17 ASDASVADVKRIETTGQSTYRADQOQMLIYQKILKDETTLESNGVAENSLFLVIMLSK 76
Db 72 ATPEASQGSANQ-----QQSQTGAGESQPVLSSTEMATVKEETVPETK--VENNVV-----120
```

```
Qy 77 AKASSSGASTATTAKAPA--TLAQ--PAAPVAPAASVARTPTQ--APVATAETAPPSVQPOA 132
Db 121 -----QEATVTPAQVPAVENVSQPTQTVAAPAAPQPAQVAPQATAGIQQAQPPVA 174

Qy 133 APAAT-----VAATDDADVISOAASN-----LVFGNNLEQTIQIILDMGGGTWER 177
Db 175 TETATAEQPVAAATTTVEQMPQAAASPAPILETPOVMTQTAPVEETQAPVVTESPAQOP 234

Qy 178 DTVVRALRAAYNNPERAIDLYSGIPENVEAQVAPARAAGQQTNQQAASPAQPAVALPV 237
Db 235 AQV-----AAPEQPAEVAPQATAGI--QQAQOPQVAVATETATAEQPVAAATTTVEQMPQAAAE 288

Qy 238 QPSPASAGPNA-----NPLNLFPGQVPSGSGSNPGVVPAGSGGALDALR 280
Db 289 SPAPISSETPOVMTQTAPVEETQAPVVTESPAQPAQVAAPEQPAEVAPQATAGIQ--344

Qy 281 QLPQFQALLQVQANPOILOPMLQELGKQ---NPQILRLIQENQAEFLRLVNESPEGGPG 337
Db 345 -----QAQOPQVAAEAQVQPPVQTAQTRPVAQPV--VVAEAQ-----VVQPPVKAQA 391

Qy 338 GNIILGQLAAAVPQTLTVPREAIQRL 366
Db 392 AQPVVKDQAQP--VASVAPQATAGIQQAQ 419

RESULT 13
US-10-234-432-59
; Sequence 59, Application US/10234432
; Publication No. US20030091598A1
; GENERAL INFORMATION:
; APPLICANT: Homer, Mary J.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Persing, David H.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF BABESIA INFECTION
; FILE REFERENCE: 210121.560
; CURRENT APPLICATION NUMBER: US/10/234,432
; CURRENT FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 59
; TYPE: PRT
; ORGANISM: Babesia sp. WA1
US-10-234-432-59

Query Match          6.9%; Score 141.5; DB 15; Length 638;
Best Local Similarity 24.2%; Pred. No. 0.023;
Matches 89; Conservative 39; Mismatches 119; Indels 121; Gaps 17;

Qy 17 ASDASVADVKRIETTGQSTYRADQOQMLIYQKILKDETTLESNGVAENSLFLVIMLSK 76
Db 72 ATPEASQGSANQ-----QQSQTGAGESQPVLSSTEMATVKEETVPETK--VENNVV-----120

Qy 77 AKASSSGASTATTAKAPA--TLAQ--PAAPVAPAASVARTPTQ--APVATAETAPPSVQPOA 132
Db 121 -----QEATVTPAQVPAVENVSQPTQTVAAPAAPQPAQVAPQATAGIQ--QAQ 170

Qy 133 APAATVAATDDADVISOAASNLVFGNNLEQTIQIILDMGGGTWERDITVVRALRAAYNNPE 192
Db 171 QPVATETAEQPV--AATT-----TEVQMPQAAAESPA 202

Qy 193 RAID-----YLYSGIPENVEAQVAPARAAGQQTNQQAASPAQPAVALP-----VQPS 240
Db 203 PILETPOVMTQTAPVEETQAPVVTESPAQPA--QVAAPEQPAEVAPQATAGIQQAQ 260

Qy 241 PASA-----GPNANPLNLFPGQ-----VPSGSGSNP-----GVVP 269
Db 261 PVAAEAQVQPPVQTAQTRPVAQPVVVAEAQVQPPVKAQAQPVVKDQAQPVASVAP 320

Qy 270 GAGSGALDALRQLPQFQALLQVQANPQILQPM-----QELGKQNPQILR 315
```

Db 321 QATAGIQ-----QAQPPVAAEAQVPPVKAARKPKPIVKDQAQPVAPVAPQATA 371  
Qy 316 LIQENQAE 323  
Db 372 GVAEDQSE 379

RESULT 14  
US-10-146-473-50  
; Sequence 50, Application US/10146473  
; Publication No. US2003010888A1  
; GENERAL INFORMATION:  
; APPLICANT: Scanlan, Matthew  
; APPLICANT: Gout, Ivan  
; APPLICANT: Stockert, Elisebeth  
; APPLICANT: Gure, Ali  
; APPLICANT: Chen, Yao-Tseng  
; APPLICANT: Old, Lloyd  
; TITLE OF INVENTION: Breast Cancer Antigens  
; FILE REFERENCE: L00461/70130(JRV)  
; CURRENT APPLICATION NUMBER: US/10/146,473  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/291,150  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 50  
; LENGTH: 2971  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-146-473-50

Query Match 6.9%; Score 141.5; DB 15; Length 2971;  
Best Local Similarity 25.5%; Pred. No. 0.18;  
Matches 95; Conservative 25; Mismatches 142; Indels 111; Gaps 16;  
Qy 79 ASSSGASTATTAKAPAT--LAQP-AAPV-----APAASVARTPTQAPVATAETAPPSVOP 130  
Db 1177 SSSLFISVPTTLPAASAPLTIPISAPLTVSASGALLTSVTPPLAPVVPAPGPPSLQP 1236  
Qy 131 QAAPAAATVAATDDADVISOAASNLVFGNN--LEOTIQOILDMGGGTWERDVTVRAALRAAY 188  
Db 1237 SCASPSASALTGLATAPSLSSSTQPGHLLLATSSHVPL-----NSTVAPACSPVL 1290  
Qy 189 NNPERAIDYLSGIPENVEAQVAPAPA-----AGOOTNQQAASPAQAVA 234  
Db 1291 -VPASALASPPPSAPNPAPQAASLLAPASASQALATPLAPMAAPQTAILAPSPAPPLAP 1349  
Qy 235 LPV-OPSPASA-----GPNANP-----LNLFPQGV 258  
Db 1350 LPVLAPSGAARVLAASSQTVPVWAPSSPTGTSLASASPVAPTPVLPASSTQTMLPAPV 1409  
Qy 259 PS-----GGSNPGVVPVGGAGDALRLQLPQFQALILQVQANPQ 297  
Db 1410 PSPLPSPASTQTALAPALAPTLLGSSPSQTLSTLGN-----PQGFPTQTLSTLPASSL 1465  
Qy 298 ILQP-MLQELGKNP-----QILRLIQENQAEFLRLVNESPEG-----GPGNIIQGO 343  
Db 1466 VTPAQTLSLAPGPPGLGPTQLSL-----APAPPLAPASVPGPAPAHTLTLAPASSASL 1520  
Qy 344 LAAAVPQTILTVP 356  
Db 1521 LAPASVQTLTLP 1533

RESULT 15  
US-10-109-886-8  
; Sequence 8, Application US/10109886  
; Publication No. US20020119499A1  
; GENERAL INFORMATION:  
; APPLICANT: TANABE SEIYAKU CO. LTD.  
; APPLICANT: TANIGUCHI, Tomoyasu

; APPLICANT: MIZUKAMI, Junko  
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND  
; FILE OF INVENTION: ANTAGONIST TO PPAR  
; FILE REFERENCE: TANIGUCHI-6  
; CURRENT APPLICATION NUMBER: US/10/109,886  
; CURRENT FILING DATE: 2002-04-01  
; PRIOR APPLICATION NUMBER: 09/514,247  
; PRIOR FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/03734  
; PRIOR FILING DATE: 1998-08-24  
; PRIOR APPLICATION NUMBER: JP231084/1997  
; PRIOR FILING DATE: 1997-08-27  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 2441  
; TYPE: PRT  
; ORGANISM: mouse  
US-10-109-886-8  
Query Match 6.9%; Score 140; DB 14; Length 2441;  
Best Local Similarity 21.7%; Pred. No. 0.18; Indels 132; Gaps 18;  
Matches 92; Conservative 41; Mismatches 159;  
Qy 44 QMLIYQKILKDE--TTLESNGVAENSLVIMLSKAKASSGASTATTAKAPATILAQPAAP 102  
Db 1857 QHCLQQAQLMERRMATNTRNPQSL-----PSPTSAPPGTPTQOESTPTQTPPP 1907  
Qy 103 VAPAS-----VART--PT-----QAPVATATAPSVQVQAPAAATVAATDDAD 145  
Db 1908 AQQPSPVMSPPAGFPNVARTQPTIVSAGKTQVPAAPPAPQF---PPAAVZAARQIE 1964  
Qy 146 VYQAASNLVFGNNLEOTIQOILDMGGGTWERDVTVRAALRAAYNNPERAIDYLSGIEP- 204  
Db 1965 REAQOQHLYRAN-----INNGMPPG-----RDGMGTGSGMTFVGLNVPRP 2006  
Qy 205 NVEAQPVARAPAAQQTN---QQAASPAQPAVALPVQSPSPASAGPNANPLNLPFGQVPS 260  
Db 2007 NOVSGPVMSMPPQWQQAQPIQOQPMGMPRPVMSQAQAAVAGPR-----MPN 2056  
Qy 261 GGSNPGVVPVGGAGSGLDALRQL-----PQFALLQLVQANPQIL----- 299  
Db 2057 VQPNRSISP---SALQDLLRLTKSPSPSQOQQVNLKSNPQLMAAFIKORTAKYVANQ 2113  
Qy 300 -----QPMQLGKQNPQ-----ILRLIQENQAEFLRLVNESPEGPGG-NILQOLAA 346  
Db 2114 PGMQPPGLSQPGMGPQPGMHQOQPSLQNLNMAQAGVPRPGVPPPPQPMGGLNPGQALN 2173  
Qy 347 AV-----PQTLTVPPEERAIQRLGGMGNRELVEVFFACNKBELTANTYLLDHGHEFD 401  
Db 2174 IMNFGHNPMTNMPQYREWVR-----QLLQHQOQOQ 2206  
Qy 402 DQQQ 405  
Db 2207 QQQQ 2210  
Search completed: December 17, 2003, 06:31:55  
Job time : 139.842 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:16:40 ; Search time 14.6701 Seconds  
(without alignments)  
1168.082 Million cell updates/sec

Title: US-09-805-550-2

Perfect score: 2036

Sequence: 1 MKNVKTLLKGNFIEASPD.....EELTANYLLDHGHFDDQQQ 405

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2036	100.0	405	US-09-413-574-2
2	900.5	44.2	368	US-09-413-574-4
3	158.5	7.8	126	US-08-747-562-26
4	151	7.4	79	US-09-100-802-4
5	150.5	7.4	8991	US-08-714-741-32
6	141.5	6.9	2372	US-09-579-181-2
7	141.5	6.9	3118	US-09-579-181-1
8	140	6.9	2441	US-08-194-468-2
9	140	6.9	2441	US-08-961-739-2
10	140	6.9	2441	US-09-514-247A-8
11	139	6.8	551	US-09-252-991A-32760
12	138	6.8	81	US-09-100-802-5
13	138	6.8	1190	US-09-107-532A-7146
14	135.5	6.7	915	US-08-480-917-2
15	135.5	6.7	915	US-09-138-736-2
16	135.5	6.7	915	US-08-988-243-2
17	135	6.6	399	US-09-252-991A-22853
18	134.5	6.6	447	US-08-450-360-4
19	134.5	6.6	486	US-08-450-360-2
20	134.5	6.6	664	US-09-328-352-6193
21	134	6.6	1088	US-09-233-857-13
22	133.5	6.6	805	US-09-103-429A-4
23	133.5	6.6	2442	US-09-514-247A-10
24	133	6.5	571	US-09-252-991A-30533
25	133	6.5	894	US-09-854-856-54
26	133	6.5	954	US-09-854-856-2
27	133	6.5	2157	US-09-854-856-52

#### ALIGNMENTS

##### RESULT 1

US-09-413-574-2  
; Sequence 2, Application US/09413574  
; Patent No. 6235972

; GENERAL INFORMATION:

; APPLICANT: Mahajan, Pramod B.

; APPLICANT: Tagliani, Laura

; TITLE OF INVENTION: Maize Rad23 Genes and Uses Thereof

; FILE REFERENCE: 0964

; CURRENT APPLICATION NUMBER: US/09/413,574

; EARLIER FILING DATE: 1999-10-06

; EARLIER APPLICATION NUMBER: 60/109,728

; EARLIER FILING DATE: 1998-11-23

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 405

; TYPE: PRT

; ORGANISM: Zea mays

US-09-413-574-2

Query Match	100.0%	Score 2036;	DB 3;	Length 405;
Best Local Similarity	100.0%	Pred. No. 7e-166;		
Matches 405;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKNVKTLLKGNFIEASPDASVADVKRIIETTTQGSTYRADQQMLIYQKILKDETTLE	60	
Db	1	MKNVKTLLKGNFIEASPDASVADVKRIIETTTQGSTYRADQQMLIYQKILKDETTLE	60	
Qy	61	SNVAENSLVIMLSKAKASSGASTATTAKAPATLAQPAAPVAPAAASVARTPTQAPVAT	120	
Db	61	SNVAENSLVIMLSKAKASSGASTATTAKAPATLAQPAAPVAPAAASVARTPTQAPVAT	120	
Qy	121	AETAPPSVQPOAAAPAAATVAAATDDADVYSQAASNLVFGNNLEQTQQIILDMGGTWERDVT	180	
Db	121	AETAPPSVQPOAAAPAAATVAAATDDADVYSQAASNLVFGNNLEQTQQIILDMGGTWERDVT	180	
Qy	181	VRALRAAYNNFERAIDLYSGIPENVEAQPVARAPAAQQTNOQAASPAQPAVALPVQPS	240	
Db	181	VRALRAAYNNFERAIDLYSGIPENVEAQPVARAPAAQQTNOQAASPAQPAVALPVQPS	240	
Qy	241	PASAGPNANPLNLPQGVPSGSGNPVPGAGSGALDRLQIPQFALLQLVQANPQILQ	300	
Db	241	PASAGPNANPLNLPQGVPSGSGNPVPGAGSGALDRLQIPQFALLQLVQANPQILQ	300	
Qy	301	PMQLQELGKQNPQILRLIQENQAEFLRLVNESPEGPGGNNILGQLAAAVPQTTLVTPEERE	360	
Db	301	PMQLQELGKQNPQILRLIQENQAEFLRLVNESPEGPGGNNILGQLAAAVPQTTLVTPEERE	360	
Qy	361	ATQRLGEGMGNRELVLVFFACNDELTANYLLDHGHEFDQQQ	405	

Db 361 A I Q R L E G M G F N R E L V L E V F F A C K D E E L T A N Y L L D H G H E F D D Q Q 405

## RESULT 2

```

RESULT 2
US-09-413-574-4
; Sequence 4, Application US/09413574
; Patent No. 6235972
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Taghiani, Laura
; TITLE OF INVENTION: Maize Rad23 Genes and Uses Thereof
; FILE REFERENCE: 0964
; CURRENT APPLICATION NUMBER: US/09/413,574
; CURRENT FILING DATE: 1999-10-06
; EARLIER APPLICATION NUMBER: 60/109,728
; EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-09-413-574-4

```

Query Match 44.2%; Score 900.5; DB 3; Length 368;  
Best Local Similarity 49.0%; Pred. No. 5.9e-69;  
Matches 197; Conservative 53; Mismatches 117; Indels 35

[illegible]

RESULT 3

```

RESULT 3
US-08-747-562-26
; Sequence 26, Application US/08747562
; Patent No. 6579697
; GENERAL INFORMATION:
; APPLICANT: WALLACH, David
; APPLICANT: BOLDIN, Mark
; APPLICANT: METT, Igor
; APPLICANT: VARFOLOMEEV, Eugene
; TITLE OF INVENTION: MODULATOR OF TNF/NGF SUPERFAMILY RECEPTORS
; NUMBER OF INVENTION: AND SOLUBLE OLIGOMERIC TNF/NGF SUPERFAMILY RECEPTORS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
;

```

```

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORMAT:
MEDIUM TYPE: FLOPPY
COMPUTER: IBM PC
OPERATING SYSTEM:
SOFTWARE: Patent
CURRENT APPLICATION NUMBER:
APPLICATION NUMBER:
PRIOR APPLICATION DATE:
APPLICATION NUMBER:
FILING DATE: 11-1-87
PRIOR APPLICATION DATE:
APPLICATION NUMBER:
FILING DATE: 11-1-87
PRIOR APPLICATION DATE:
APPLICATION NUMBER:
FILING DATE: 02-6-87
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, ROSEN
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-666-7777
TELEFAX: 202-737-7777
INFORMATION FOR SEQ ID:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-747-562-26

```

Query Match 7.8%; Score 158.5; DB 4; Length 126;  
Best Local Similarity 31.1%; Pred. No. 3e-06;  
Matches 38: Conservative 25; Mismatches 50; Indels 9

Qy	13	FEIASPDPASVADVKKRIITETQGOSTYRADQOMLIYQCGKILKDTELTLESNGVAENSFLVI	72
		: :	
Db	2	KFIRMEPDVTVKLKEIEAEKGRDAPFPAQCKLIYAGKILSDDDVPIDRYDIRIDEKNFVVV	61
		: :	
Qy	73	MUSKAKA-----SSSGASTATTAKAPATLAQPAAPVA-----PAASVARTPTQAPVATAET	120
		: :	
Db	62	VMTKTGAGGTSAPEASPTAAPESSFTFPAPTSGMSPHPFAAREDKSPSEESTPTPEQ	12
		: :	
Qy	124	AP 125	
Db	122	EP 123	

**RESULT 4**

```

RESULT 4
US-09-100-802-4
; Sequence 4, Application US/09100802A
; Patent No. 6294363
; GENERAL INFORMATION:
; APPLICANT: Madura, Kiran
; TITLE OF INVENTION: Methods and Compositions for the Rapid
; TITLE OF INVENTION: Purification of Proteasomes and Methods of Use of Components
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: UMDNJ97-11
; CURRENT APPLICATION NUMBER: US/09/100,802A
; EARLIER APPLICATION NUMBER: 1998-06-19
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Artificial Sequence

```



```

Query Match      7.48;  Score 150.5;  DB 4;  Length 8991;
Best Local Similarity 21.2%;  Pred. No. 0.0072;
Matches 8;  Conservative 52;  Mismatches 152;  Indels 109;  Gaps 15;

QY 20 DASVADVVKRIETQ-----GOSTYRQQQMLLYQGKILKDTTLESNGVAENSFLVI 72
   |||:::|
Db 5184 DASIAKLEKVEQFQNSGDGEQAGYLAAGEDLIKAELAEKARADLK-KAVDPE---- 5238

```

Query Match	6.9%;	Score	141.5;	DB	4;	Length	2972;
Best Local Similarity	25.5%;	Pred.	No.	0.0085;			
Matches	95;	Conservative	25;	Mismatches	142;	Indels	111;
Gaps	16;						
QY	79	ASSSGASTATTAKAPAT--LAQP-AAPV-----	APASVARIT	TQ	PVAT	AEAT	PPSVQP
DB	1177	SSSLPISVPTTLAPAPASALPTIPISAPLTVSASGPALLTSVTPLPLVVPVPAAPGPPSLQP					1236
QY	131	QAAPAAATVAATDDADVYQAASNLVGN--LQETIQILLDMGGTWERDTVVRAALRAY	188				
DB	1237	SGASPSASALTGLATAPASLSSQTPGCHPLLPAPTSSHVPGL-----	NTVPAPACSPVL	1290			
QY	189	NNPERAIDVLYSGIPENVEAQPVARAPA-----	AGQQTNOQAASPAQPAVA	234			
DB	1291	-VPASLASPPFSAPNPAQAASLLAPASASOALATPLAPMAAPQTAILAPSPAPPLAP	1349				
QY	235	LPV-QPSPASA-----	GNANP-----	LMLFPQGV	258		
DB	1350	LPVLASPGNAVPLASSQTPVPVMASSIPGTSLASASVPAPTPVLPAPSSQTQMLPAPV	1409				
QY	259	PS-----	GGSNPGVVPAGSGALDALKQLFOFALLQIVQANPQ	297			
DB	1410	PSPLSPASQTLLALAPALAPTUGGSSPSQTLSLGTGN-----	PQGPPTQTLSTLTPASSL	1465			
QY	298	ILQP-MLQELGKNP---QILRLIQENQAEFLRLVNESPEG-----	GGCGNITLQG	343			

Db 1466 VPTPAQTLTSLAPGPGTQTLSL-----APAPPLAPASVPGPAPARTLTLPASSASL 1520  
Qy 344 LAAAVPQTLTVP 356  
Db 1521 LAPASVQTLTSLP 1533

## RESULT 7

US-09-579-181-1  
; Sequence 1, Application US/09579181  
; Patent No. 6365372  
; GENERAL INFORMATION:  
; APPLICANT: Chiviva, John  
; APPLICANT: Yaciuk, Peter  
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)  
; FILE REFERENCE: 16153-4247  
; CURRENT APPLICATION NUMBER: US/09/579,181  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: 60/136,620  
; PRIOR FILING DATE: 1999-05-27  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 3118  
; TYPE: PRT  
; ORGANISM: Human  
US-09-579-181-1

Query Match 6.9%; Score 141.5; DB 4; Length 3118;  
Best Local Similarity 25.5%; Pred. No. 0.0091;  
Matches 95; Conservative 25; Mismatches 142; Indels 111; Gaps 16;

Qy 79 ASSSGSTATTAKPAT--LAQP-AAPV-----APAAVARTTQAPVATAETAPPSVOP 130  
Db 1323 SSSLPTISVPTTLPAPASAPLTIPISAPLTVSASGPALLTSVTPPLAPVVPVAPGPPSLOP 1382  
Qy 131 QAAPAAVTAATDDADVYQAASNLVFGNN--LQTTIQIILDMGGGTWERTVVRALRAAY 188  
Db 1383 SGASPSALTLGLATAPSLSSSQTPGHPLLLAPTSSHPGL-----NSTVAPACSPVL 1436  
Qy 189 NNPERAIDLYSGIPENVEVAQVARAPA-----AGQTNQOAAAPAPAVA 234  
Db 1437 -VPASALASPPSAEPAPAPQAQALLAPASSAQALATPLAPMAAPQTAILAPSPAPPLAP 1495  
Qy 235 LPV-QPSPASA-----GPNAP-----LNLPQGV 258  
Db 1496 LPVLAPSPGAAPVLAASQTPVPVMAFSSTEGTSLASASVPVAPTPVLAESPSTQMLPAPV 1555  
Qy 259 PS-----GGSNPVGVPAGSGALDALRQLPQFQALLQVQNPQ 297  
Db 1556 PSPLPSPASTQTLALAPALAPTLLGGSPSQTLISLTGN-----PQGPFTQTLSTPASSL 1611  
Qy 298 ILQP-MLQELGKNP-----QILRLIQENQAEFLRLVNESPEG-----GPGNILGQ 343  
Db 1612 VPTPAQTLTSLAPGPGTQTLSL-----APAPPLAPASVPGPAPARTLTLPASSASL 1666  
Qy 344 LAAAVPQTLTVP 356  
Db 1667 LAPASVQTLTSLP 1679

## RESULT 8

US-08-194-468-2  
; Sequence 2, Application US/08194468  
; Patent No. 5750336  
; GENERAL INFORMATION:  
; APPLICANT: Montminy, Marc R.

; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF cAMP AND MITOGEN  
; TITLE OF INVENTION: RESPONSE GENES  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark

; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/194,468  
; FILING DATE: 10-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: P41 9672  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619)-546-4737  
; TELEFAX: (619)-546-9392  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2441 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-194-468-2

Query Match 6.9%; Score 140; DB 1; Length 2441;  
Best Local Similarity 21.7%; Pred. No. 0.0085;  
Matches 92; Conservative 41; Mismatches 159; Indels 132; Gaps 18;

Qy 44 QMLYIQGKILKDE--TTLESNGVAENSLVIMLSKAKASSSGASTATTAKAPATLAQPAAP 102  
Db 1857 QHCLQAQALMRRMATNRNVPOQSL-----PSPTSAPPGTPTQQPSTPQTPOPP 1907  
Qy 103 VAPAAAS-----VART--PT-----QAPVATAETAPPSVQQAAPAAVTAATDDAD 145  
Db 1908 AQQPSPVNMSPAGFPNVARTQPTIVSAGKTNQVPAPPPAQP---PPAAVEAARQIE 1964  
Qy 146 VYSQAASNLVFGNNLEQTTIQIILDMGGGTWERTVVRALRAAYNNPERAIDLYSGIPE- 204  
Db 1965 REAQQQOHLRYAN-----INNGMPG-----RDGMGTGSGMTPEVGLNVPRP 2006  
Qy 205 NYEAPQVAPAPAGQOTN-----QQAASPAQPAVALPVQSPASAGPNANPLNLPQGVPS 260  
Db 2007 NQVSGFVMSMPPGQWQAPIPQQQPMFMPRVMSMQAQAAGVGR-----MPN 2056  
Qy 261 GGSNFGVWPGAGSGALDALROL-----PQFQALLQVQNPQIL----- 299  
Db 2057 VQPNRSISP---SALQDLRLTLKSPSSPQQQQOVLNLSKNPQLMAAFIKQRTAKTVANQ 2113  
Qy 300 -----QPMQLQELGKNPQ-----ILRLIQENQAEFLRLVNESPEGGPGG-NILGQLAA 346  
Db 2114 PGWQPGQGLSQPGQPGQPMHQQPSLQNLNAMQAGVPRPGVPPPPQPMAGMLNPPQOALN 2173  
Qy 347 AV-----PQTLTVTPEERAIQRLSGMGNRELVLVFPACNKDBELTANYLLDGHGEPD 401  
Db 2174 INNPNGHNPMNTNMNPQYREVMVR-----QLLQHQOQOQ 2206  
Qy 402 DQOQ 405  
Db 2207 QOQO 2210

## RESULT 9

US-08-961-739-2  
; Sequence 2, Application US/08961739A  
; Patent No. 6063583  
; GENERAL INFORMATION:  
; APPLICANT: Montminy, Marc R.  
; TITLE OF INVENTION: Methods for Treating Diabetes Mellitus

FILE REFERENCE: SALK1650-1  
CURRENT APPLICATION NUMBER: US/08/961,739A  
CURRENT FILING DATE: 1997-10-31  
EARLIER APPLICATION NUMBER: US 194,468  
EARLIER FILING DATE: 1994-02-10  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 2441  
TYPE: PRT  
ORGANISM: Mus  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)...(2441)  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-08-961-739-2

Query Match 6.9%; Score 140; DB 3; Length 2441;  
Best Local Similarity 21.7%; Pred. No. 0.0085;  
Matches 92; Conservative 41; Mismatches 159; Indels 132; Gaps 18;  
QY 44 QMLIYQGKILKDE--TTLESNGVAENSFLVIMLSKAKASSSGASTATTAKAPATLAQPAAP 102  
DB 1857 QHCLQQAQLMRRMATMTRNVFQOQL-----PSPTSAPPCTPTQQPSTPTQPP 1907  
QY 103 VAPAAAS-----VART--PT---QAPVATAETAPPSVQPAAPAAATVAATDDAD 145  
DB 1908 AQOPSPVNMSPAGFPNVARTQPTIVSAGKPTNQVPPAPQ---PPAAVEAARQIE 1964  
QY 146 VYQAASNLVFGNNLEQTIQIILDMGGTWTERTVVRALRAAYNNPRAIDLYSGIPE- 204  
DB 1965 REAQOQHLYRAN-----INNGMPG-----RDGMGTGSGQMTFVGLNVPRP 2006  
QY 205 NVEAQPVARAPAAAGQTN---QQAASPAQPAVALPVQPSPASAGPNANPLNLPQGVPS 260  
DB 2007 NOVSGPVNMSPGQWQOQAPIPQOPMGPMPRPVMSMQAAVAGPR-----MPN 2056  
QY 261 GGSNPGVVPAGSGALDALRLQ-----POFALLQLVQANPQIL----- 299  
DB 2057 VQPNRSISP---SALQDLRLTKSPSPQOQQVNLILKSNPQLMAAFIKORTAKYVANO 2113  
QY 300 -----QPMQLGKQNPQ-----ILRLIOENQAELRLVNESPEGGPGG-NILGQLAA 346  
DB 2114 PGMQOPGLOSPGMQOPGMHQPSLQNLNAMQAGVPRPGVPPQPMGGLNPQGOALN 2173  
QY 347 AV-----PQTLTVTPEREAIQRLGEGFNLVLEFFACNKKDEELTANYLLDHGHEFD 401  
DB 2174 IMNPGHNPNMTNMNPQYREMYRR-----QLLQHQOQQQ 2206  
QY 402 DQQQ 405  
DB 2207 QQQQ 2210

RESULT 10  
US-09-514-247A-8  
Sequence 8, Application US/09514247A  
Patent No. 6365361  
GENERAL INFORMATION:  
APPLICANT: TANABE SEIYAKU CO. LTD.  
APPLICANT: TANIGUCHI, Tomoyasu  
APPLICANT: MIZUKAMI, Junko  
TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO PPA  
FILE REFERENCE: TANIGUCHI=6  
CURRENT APPLICATION NUMBER: US/09/514,247A  
CURRENT FILING DATE: 2000-02-28  
PRIOR APPLICATION NUMBER: PCT/JP98/03734  
PRIOR FILING DATE: 1998-08-24  
PRIOR APPLICATION NUMBER: JP231084/1997  
PRIOR FILING DATE: 1997-08-27  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 8

LENGTH: 2441  
TYPE: PRT  
ORGANISM: mouse  
US-09-514-247A-8

Query Match 6.9%; Score 140; DB 4; Length 2441;  
Best Local Similarity 21.7%; Pred. No. 0.0085;  
Matches 92; Conservative 41; Mismatches 159; Indels 132; Gaps 18;  
QY 44 QMLIYQGKILKDE--TTLESNGVAENSFLVIMLSKAKASSSGASTATTAKAPATLAQPAAP 102  
DB 1857 QHCLQQAQLMRRMATMTRNVFQOQL-----PSPTSAPPCTPTQQPSTPTQPP 1907  
QY 103 VAPAAAS-----VART--PT---QAPVATAETAPPSVQPAAPAAATVAATDDAD 145  
DB 1908 AQOPSPVNMSPAGFPNVARTQPTIVSAGKPTNQVPPAPQ---PPAAVEAARQIE 1964  
QY 146 VYQAASNLVFGNNLEQTIQIILDMGGTWTERTVVRALRAAYNNPRAIDLYSGIPE- 204  
DB 1965 REAQOQHLYRAN-----INNGMPG-----RDGMGTGSGQMTFVGLNVPRP 2006  
QY 205 NVEAQPVARAPAAAGQTN---QQAASPAQPAVALPVQPSPASAGPNANPLNLPQGVPS 260  
DB 2007 NOVSGPVNMSPGQWQOQAPIPQOPMGPMPRPVMSMQAAVAGPR-----MPN 2056  
QY 261 GGSNPGVVPAGSGALDALRLQ-----POFALLQLVQANPQIL----- 299  
DB 2057 VQPNRSISP---SALQDLRLTKSPSPQOQQVNLILKSNPQLMAAFIKORTAKYVANO 2113  
QY 300 -----QPMQLGKQNPQ-----ILRLIOENQAELRLVNESPEGGPGG-NILGQLAA 346  
DB 2114 PGMQOPGLOSPGMQOPGMHQPSLQNLNAMQAGVPRPGVPPQPMGGLNPQGOALN 2173  
QY 347 AV-----PQTLTVTPEREAIQRLGEGFNLVLEFFACNKKDEELTANYLLDHGHEFD 401  
DB 2174 IMNPGHNPNMTNMNPQYREMYRR-----QLLQHQOQQQ 2206  
QY 402 DQQQ 405  
DB 2207 QQQQ 2210

RESULT 11  
US-09-252-991A-32760  
Sequence 32760, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 32760  
LENGTH: 551  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32760

Query Match 6.8%; Score 139; DB 4; Length 551;  
Best Local Similarity 22.1%; Pred. No. 0.0012;  
Matches 90; Conservative 58; Mismatches 129; Indels 130; Gaps 20;  
QY 20 DASVADVKKRIETQCG-----STYRADOQQLIYQKILKDETTLENGVAENS 68  
DB 2 ERNVSELIRVPDIGNGEVIELLVKPGDKVEADQSLTLE----SKASME---IPSPK 54  
QY 69 FLVIMLSKAKASSS-----GASTATTAKAPATLAQPAAPVAPAAASVARTPTQ 115

Db 55 AGVTSIKAKVGTLLKEGDEILLEVEGEQPAEAKAERAPQAPEAPKAE-----P 106  
Qy 116 APVATETAPPVQPOQAAPAAATVAATDDADVYSQAASNLVFGNNLEQTTQIILDMGGGTW 175  
Db 107 AP-----APSESKAAPAAASVQDIKVPDIGSAGKANVI-----EYVVRAGDVT 150  
Qy 176 ERDVTVRAL--RAAYNNPERAIDYLYSGIPENV-----EAQPVARAPAA 217  
Db 151 EADQSLLTLESKASMEISPA-----SGVSVSIVKGVDEVTGDLILKLVKEGAAPAA 205  
Qy 218 GOOTNOQAASPAQPAVALVQVPSASAGNPNLNLFPQGVPSGSGNPGVWPGAGSGALD 277  
Db 206 EE---QPAAPAAQ-AAAPAEQKPAAP-----APAKADTPAPV---GAPSRD 247  
Qy 278 ALRQLPQFALLQLVQANPQILQPMLOELGKONPQIL-----RLIQENQAEFLRLVNES 331  
Db 248 GAK-----VHAGPAV-RMLAREFVSELSEVKASGPKGRILKEDVQVFKVKEQLQR 295  
Qy 332 PEGGPGGNILGQLAAAVPQTLVTPE-----EREAIQRLGGMG 369  
Db 296 AKSGGAGATGG---AGIPP-----IPEVDFSKFGEVEEVANTRLMQVG 335

## RESULT 12

US-09-100-802-5  
; Sequence 5, Application US/09100802A  
; Patent No. 6294363  
; GENERAL INFORMATION:  
; APPLICANT: Madura, Kiran  
; TITLE OF INVENTION: Methods and Compositions for the Rapid  
; TITLE OF INVENTION: Purification of Proteasomes and Methods of Use of Components  
; TITLE OF INVENTION: Thereof  
; FILE REFERENCE: UMDNJ97-11  
; CURRENT APPLICATION NUMBER: US/09/100,802A  
; CURRENT FILING DATE: 1998-06-19  
; EARLIER APPLICATION NUMBER: 60/050,171  
; EARLIER FILING DATE:  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 81  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic Sequence  
US-09-100-802-5

Query Match 6.8%; Score 138; DB 3; Length 81;  
Best Local Similarity 36.4%; Pred. No. 8.9e-05;  
Matches 28; Conservative 17; Mismatches 32; Indels 0; Gaps 0;

Qy 3 LNVTKLGNFTEIASPDASVADVKRIETQSQSTYRADQQMLIYQGIKDKETTESN 62  
Db 5 ITLTKLQQQTFKIRMEPDETVKLVKEIBAERDAPFVAGKLIYAGKILSDDDVPIRDY 64  
Qy 63 GVAENSLYMLSKAKA 79  
Db 65 RIDERKNFVVMVTKTKA 81

## RESULT 13

US-09-107-532A-7146  
; Sequence 7146, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street

CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 7146:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1190 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...1190  
SEQUENCE DESCRIPTION: SEQ ID NO: 7146:  
US-09-107-532A-7146

Query Match 6.8%; Score 138; DB 4; Length 1190;  
Best Local Similarity 24.6%; Pred. No. 0.0044;  
Matches 69; Conservative 29; Mismatches 112; Indels 70; Gaps 13;

Qy 75 SKAKASSGASTATTAKAPATLAOPA-----APVAPASVARTP--T 114  
Db 133 SARQTAEPATGLAAAKALKSRKKOPAPREDADGQTLDMFGLSSEPEQPAATPEPEPELS 192  
Qy 115 QAPVATAETAPPSVQ-----PQAAP-----AATVAATDDADVYSQAASNLVFGNNLEQTTQ 165  
Db 193 QAETGASETAVPILASKPEPPQETPNIPGAAPIMAESPADRYDLGYGHM--GNGL--TW 248  
Qy 166 QILDMGGGTWERDVTVRALRAAYNNPERAIDYLYSGIPENV--EAQPVARAP-----AQQ 219  
Db 249 NRLEBEHDYK-----TVAHTAPDRVTTFYDADMPPEIREKTIQKVAATAEMISATQ 300  
Qy 220 OTNQQAASPAQPAVALP-----VOPSPASAG--PNANPLNFPQGVPSGSGNPGVV 268  
Db 301 DTPVFSTPQPERVQPGNSEPEPEKVDAPAAVNPEPETGDSNTPVPSPAQKAGAP----- 357  
Qy 269 PGAGS-----GALDALRQLPQFQALLQLVQANPQIL 299  
Db 358 TGTGSPQTQEAAPKTAADGLNLTPNVEEVLNKAQYPODKL 397

## RESULT 14

US-08-480-917-2  
; Sequence 2, Application US/08480917  
; Patent No. 5820864  
; GENERAL INFORMATION:  
; APPLICANT: PARANHOS-BACCALA, Glaucia  
; APPLICANT: LESENECHAL, Mylene  
; APPLICANT: JOLIVET, Michel  
; TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE

CORRESPONDENCE ADDRESS:  
ADDRESS: Oliff & Berridge  
STREET: 700 South Washington Street, Suite 300  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/138,736  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,917  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Berridge, William P.  
REGISTRATION NUMBER: 30,024  
REFERENCE/DOCKET NUMBER: WPB 36400  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6400  
TELEFAX: 703-836-2787  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 915 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "protein"  
US-09-138-736-2

Query Match	6.7%;	Score 135.5;	DB 3;	Length 915;					
Best Local Similarity	23.1%;	Pred. No. 0.005;							
Matches	63;	Conservative 43;	Mismatches 126;	Indels 41; Gaps 8;					
QY	71	VTMLS	KAKASSG	STATTAKAPATLAQAPAA	PVAPAA	SVARPTQ	PVATAETAP	SVQP	130
DB		:	:	:	:	:	:	:	
	369	VVATAK	PGCVSSG	TDAASSHTTT	TSAAA	ASP-----	PVSAPAKAA	--APP	AAAR 419
QY	131	QAAPAA	TVAATD	DADV	SQAASNL	VFGNN	LEOTIQ	ILLDMGGT	WEDTVV
DB		:	:	:	:	:	:	:	
	420	SAEP-----	HVGS	KIIAN	LVGIN	QTVRS	WSTG	STGAPAT	TRSTAVTSTTTA----
QY	191	PERAID	LYSGIP	-----	ENVEA	QPVARA	PAAG	QTNQQAAS	PAQPAVAL
DB		:	:	:	:	:	:	:	
	467	PORTSP	YHNGRP	VTAGL	VAANS	GSA	SSPTAAK	PTGEE	KASAA-----
QY	244	AGFNAN	PLMFP	QGVSP	GSGSN	GVPG	GAGS	GALDAL	RQLPQ
DB		:	:	:	:	:	:	:	
	518	VAINAT	RPAHNAS	LPOA	PTD-GV	LAAV	YQSE	GEVHQ--	SIERLES
QY	304	QELG	KQNQ	POILRL	IQENQA	BFLRV	NES	PEGGP	336
DB		:	:	:	:	:	:	:	
	575	DTIRRD	HEOLLN	LGLEA	OMTE	LOOSR	PTP	OTOP	607
DB		:	:	:	:	:	:	:	
	574								

Search completed: December 17, 2003, 06:21:21  
Job time : 15.6701 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:16:40 ; Search time 16.6624 Seconds  
(without alignments)  
2123.954 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 1910

Sequence: 1 MKLTVKTLKGTHTFEIRVQPN.....CDRNEELAANYLLEHAGEED 368

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1153.5	60.4	379	2 T14337	RAD23 protein, iso
2	1043.5	54.6	367	2 F96827	protein F20817.8
3	954	49.9	382	2 T14336	RAD23 protein, iso
4	887.5	46.5	392	2 T04150	RAD23 protein homo
5	661.5	34.6	246	2 G86296	T24D18.27 protein
6	640.5	33.5	385	2 JC7783	RAD 23B protein -
7	625.5	32.7	409	2 S44346	RAD23 protein homo
8	603.5	31.6	363	2 S44443	RAD23 protein homo
9	561.5	29.4	368	2 T40115	uv excision repair
10	424	22.2	113	2 H86296	F309.1 protein - A
11	405	21.2	372	2 T27774	hypothetical prote
12	403	21.1	398	2 S50507	excision repair pr
13	204.5	10.7	142	2 T51479	hypothetical prote
14	201	10.5	551	2 C84549	probable ubiquitin
15	195.5	10.2	536	2 B84549	probable ubiquitin
16	163.5	8.6	354	2 T38404	yeast dak2 homolog
17	163.5	8.6	502	2 T18562	hypothetical prote
18	159	8.3	373	2 S54583	ubiquitin-like pro
19	143	7.5	324	2 T02358	ubiquitin homolog
20	140	7.3	701	2 JN0874	ubiquitin-like fus
21	138	7.2	313	2 JQ1728	ubiquitin precurs
22	137.5	7.2	292	2 T18584	F15C11.2 - Caenorh
23	136	7.1	1135	2 T30561	scythe protein - A
24	135	7.1	356	1 UQUTRC	polyubiquitin / ri
25	132.5	6.9	990	2 T14756	hypothetical prote
26	130.5	6.8	76	2 S31653	ubiquitin - Glardi
27	127.5	6.7	80	2 T28305	ORF MSV144 probabl
28	126.5	6.6	77	2 S29238	ubiquitin - Entamo
29	126.5	6.6	156	1 UQHUR7	ubiquitin / riboso

30	126.5	6.6	156	1 I52328	ubiquitin / riboso
31	126.5	6.6	157	2 I50438	ubiquitin polyprot
32	126.5	6.6	167	2 I51568	polyubiquitin - Af
33	126.5	6.6	222	1 UQHV	ubiquitin precursor
34	126.5	6.6	229	1 UQHUB	polyubiquitin 3 -
35	126.5	6.6	229	1 S13928	ubiquitin precursor
36	126.5	6.6	231	1 A31560	polyubiquitin - fr
37	126.5	6.6	245	2 I45964	polyubiquitin - bo
38	126.5	6.6	255	2 A29584	ubiquitin precursor
39	126.5	6.6	305	1 S29853	polyubiquitin 4 -
40	126.5	6.6	305	1 I50437	polyubiquitin 4 -
41	126.5	6.6	305	1 S12583	polyubiquitin 4 -
42	126.5	6.6	381	2 S21083	polyubiquitin 5 -
43	126.5	6.6	685	1 UQHUC	polyubiquitin 9 -
44	126.5	6.6	810	2 S45359	polyubiquitin 10 -
45	126	6.6	155	1 UQW07A	ubiquitin / riboso

ALIGNMENTS

RESULT 1

T14337

RAD23 protein, isoform II - carrot

C;Species: Daucus carota (carrot)

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000

C;Accession: T14337

R;Sturm, A.; Leinhardt, S.

Plant J. 13, 815-821, 1998

A;Title: Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in yeast.

A;Reference number: Z17989; MUID:98345997; PMID:9681019

A;Accession: T14337

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-379 <STU>

A;Cross-references: EMBL:Y12014; NID:GI914684; PIDN:CAA72742.1; PID:GI914685

A;Experimental source: subspecies Queen Anne's Lace, isolate W001C

C;Genetics:

A;Gene: RAD23-2

C;Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match	60.4%;	Score	1153.5;	DB 2;	Length	379;			
Best Local Similarity	63.0%;	Pred. No.	5.2e-70;						
Matches	240;	Conservative	45;	Mismatches	81;	Indels	15;	Gaps	9;
QY	1	MKLTVKTLKGTHTFEIRVQPN	DTIMAVKNIIEIQKDSYPMGQQLIFNGKVLKDESTLE	60					
DB	1	MKLTVKTLKGTHTFEIRVQPN	DTIMAVKNIIEIQKDSYPMGQQLIFNGKVLKDESTLE	60					
QY	61	ENKVNEDGFLVWMLSGKTS	GTSGTSSQHSNTPATROAP---	PLEAPQAPQPPVPAPIT	117				
DB	61	ESKISDGLVWMLSGKTS	GTSGTTPAQAQSSAPATPAPAVAPAPAPAPASAVIPNT	120					
QY	118	TSQPEG--LPAQAP-NTHD	NAASNLISGRNVDTIINQLMEMGGSGWDKQVORALRAAYN	174					
DB	121	T-VPEAPLSPAPAPSDTY	GEAASNVVAGSNLEQTTQHIDMGGGMWDTNMVSRALRAAYN	179					
QY	175	NPERAVELYSGIPVTAET	AVFIG---GQAGANTDRA--PTGEAGLS-GIPNTAPIDLFP	228					
DB	180	NPERAVDYLISGIPENAE	AAVPSHFQDQIINAGNNAISDNGVAGAAPCAPNSLPLNMFP	239					
QY	229	QGASNAGGAGGGLDPL	NNPQFAVRVMVHTNPQILOPMLVELSKQNPQILRLIENH	288					
DB	240	QETLSGVTGAGLSLE	FLNNPQFQFLRSNVQRNPQILOPMLLELQKQNPQLLQOEHH	299					
QY	289	DEFLQLNPFPGEGD	FLDQPEEDEMPHAISVTPPEQBAIGRLSNGMDFRARI	348					
DB	300	EEFLQLINFEVASE	GDMFDQEQD-VQOEITVTAADQAEIERLEANGDFRGLVIEAFIA	358					
QY	349	CORNEELAANYLLE	HAGE-ED 368						
DB	359	CORNEELAVNLL	ENAGDPED 379						

A:Molecule type: mRNA  
A:Residues: 1-382 <STU>  
A:Cross-references: EMBL:Y12013; NID:g1914682; PIDN:CAA72741.1; PID:g1914683  
A:Experimental source: subspecies Queen Anne's Lace, isolate W001C  
C:Genetics:  
A:Gene: RAD23-1  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 49.9%; Score 954; DB 2; Length 382;  
Best Local Similarity 52.4%; Pred. No. 1.3e-56;  
Matches 205; Conservative 52; Mismatches 98; Indels 36; Gaps 9;

Qy 1 MKLTVKTLKGTFTFIRVQNPNDTMAVKNIEIQGKDSYPMGQQLIFNGKVLKDESTLE 60  
Db 1 MKIYVKTLGSKQFEIQVNPDDSVADVRSIETAGAAVYPAAGQMLIYQKVLKDGTTLL 60  
Qy 61 ENKVNEDGFLVWMLSKGKTSGSTSSQHSNTPATROAPPLRAPQAP-----QPVA- 114  
Db 61 ENNVAENSFVIMLSKSKSPSGSGSTTS-----TAAAPKAPQTSAPPSVPAPAVSQPPAST 116  
Qy 115 -PITTSQPEGLPAQAP-----NTHNAASNLLSGRNVDTIINQLMEMGGSGWDK 163  
Db 117 LPVPAPSPAPAPATAPIPSAAGSEANVYDSAAASLLVAGSLEGAIQIILDMGGGTWDRD 176  
Qy 164 KYQRAALAAVNNPERAVEVLYSGIPVTAETAVPI-----GGQGANTTDRAPTG--EAGLS 216  
Db 177 TVIRIVRAAFNNPERAVEVLYSGIPEQAE-APPVAPSPPSGQAANPLDQPPAAQAQAPAS 235  
Qy 217 GIPNAPLPLFPQAGSAG-AGAGGGPLDFLRNPNPQFQAVREVMHTNPQILQPLMVELSK 275  
Db 236 AGFNANPLDPLFPQGLPDMGSMAGAGNLDPLRTNQFQALRAMVQSNFQILQPLMLQELGK 295  
Qy 276 QNPQLRLIENHDFLQNLNPEPGEDEMPHAISSVTPEQEAIGRLSM 335  
Db 296 QNPHLMRLIQHQADFQLINPEMGEENLL-----GHGPOAISVTPEERDAIERLEAM 349  
Qy 336 GFDRAVIEAFLACDRNEELAANYLLEHAGE 366  
Db 350 GFDRELVLVEVFACNKEELAANYLLEHME 380

RESULT 4  
RAD23 protein homolog - rice  
C:Species: Oryza sativa (rice)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: T04150  
R:Schultz, T.F.; Quatrano, R.S.  
Plant Mol. Biol. 34, 557-562, 1997  
A:Title: Characterization and expression of a rice RAD23 gene.  
A:Reference number: 208695; MUID:97369378; PMID:9225866  
A:Accession: T04150  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-392 <SCH>  
A:Cross-references: EMBL:U63530; NID:g1488296; PIDN:AAB65841.1; PID:g1488297  
A:Experimental source: cv. Nipponbare  
C:Genetics:  
A:Gene: RAD23  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 46.5%; Score 887.5; DB 2; Length 392;  
Best Local Similarity 50.9%; Pred. No. 3.7e-52;  
Matches 199; Conservative 52; Mismatches 109; Indels 31; Gaps 10;

Qy 1 MKLTVKTLKGTFTFIRVQNPNDTMAVKNIEIQGKDSYPMGQQLIFNGKVLKDESTLE 60  
Db 1 MKISVTKLGSTFQIEVDQAQVADVVKRIETQGHYIPAEQMLIHQKVLKDDTTLD 60  
Qy 61 ENKVNEDGFLVWMLSKGKTSGSTSSQHSNTPATROAPPLRAPQAP-----QPVAPIIT 118  
Db 61 ENKVLNSFLVIMLRGKSGSSAPATSK-----APSNQAPPTQTVPAAPASQAPVAPATT 116

A:Molecule type: mRNA  
A:Residues: 1-382 <STU>  
A:Cross-references: EMBL:Y12013; NID:g1914682; PIDN:CAA72741.1; PID:g1914683  
A:Experimental source: subspecies Queen Anne's Lace, isolate W001C  
C:Genetics:  
A:Gene: RAD23-1  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 49.9%; Score 954; DB 2; Length 382;  
Best Local Similarity 52.4%; Pred. No. 1.3e-56;  
Matches 205; Conservative 52; Mismatches 98; Indels 36; Gaps 9;

Qy 1 MKLTVKTLKGTFTFIRVQNPNDTMAVKNIEIQGKDSYPMGQQLIFNGKVLKDESTLE 60  
Db 1 MKIYVKTLGSKQFEIQVNPDDSVADVRSIETAGAAVYPAAGQMLIYQKVLKDGTTLL 60  
Qy 61 ENKVNEDGFLVWMLSKGKTSGSTSSQHSNTPATROAPPLRAPQAP-----QPVA- 114  
Db 61 ENNVAENSFVIMLSKSKSPSGSGSTTS-----TAAAPKAPQTSAPPSVPAPAVSQPPAST 116  
Qy 115 -PITTSQPEGLPAQAP-----NTHNAASNLLSGRNVDTIINQLMEMGGSGWDK 163  
Db 117 LPVPAPSPAPAPATAPIPSAAGSEANVYDSAAASLLVAGSLEGAIQIILDMGGGTWDRD 176  
Qy 164 KYQRAALAAVNNPERAVEVLYSGIPVTAETAVPI-----GGQGANTTDRAPTG--EAGLS 216  
Db 177 TVIRIVRAAFNNPERAVEVLYSGIPEQAE-APPVAPSPPSGQAANPLDQPPAAQAQAPAS 235  
Qy 217 GIPNAPLPLFPQAGSAG-AGAGGGPLDFLRNPNPQFQAVREVMHTNPQILQPLMVELSK 275  
Db 236 AGFNANPLDPLFPQGLPDMGSMAGAGNLDPLRTNQFQALRAMVQSNFQILQPLMLQELGK 295  
Qy 276 QNPQLRLIENHDFLQNLNPEPGEDEMPHAISSVTPEQEAIGRLSM 335  
Db 296 QNPHLMRLIQHQADFQLINPEMGEENLL-----GHGPOAISVTPEERDAIERLEAM 349  
Qy 336 GFDRAVIEAFLACDRNEELAANYLLEHAGE 366  
Db 350 GFDRELVLVEVFACNKEELAANYLLEHME 380

RESULT 4  
RAD23 protein homolog - rice  
C:Species: Oryza sativa (rice)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: T04150  
R:Schultz, T.F.; Quatrano, R.S.  
Plant Mol. Biol. 34, 557-562, 1997  
A:Title: Characterization and expression of a rice RAD23 gene.  
A:Reference number: 208695; MUID:97369378; PMID:9225866  
A:Accession: T04150  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-392 <SCH>  
A:Cross-references: EMBL:U63530; NID:g1488296; PIDN:AAB65841.1; PID:g1488297  
A:Experimental source: cv. Nipponbare  
C:Genetics:  
A:Gene: RAD23  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 46.5%; Score 887.5; DB 2; Length 392;  
Best Local Similarity 50.9%; Pred. No. 3.7e-52;  
Matches 199; Conservative 52; Mismatches 109; Indels 31; Gaps 10;

Qy 1 MKLTVKTLKGTFTFIRVQNPNDTMAVKNIEIQGKDSYPMGQQLIFNGKVLKDESTLE 60  
Db 1 MKISVTKLGSTFQIEVDQAQVADVVKRIETQGHYIPAEQMLIHQKVLKDDTTLD 60  
Qy 61 ENKVNEDGFLVWMLSKGKTSGSTSSQHSNTPATROAPPLRAPQAP-----QPVAPIIT 118  
Db 61 ENKVLNSFLVIMLRGKSGSSAPATSK-----APSNQAPPTQTVPAAPASQAPVAPATT 116

A:Molecule type: mRNA  
A:Residues: 1-382 <STU>  
A:Cross-references: EMBL:Y12013; NID:g1914682; PIDN:CAA72741.1; PID:g1914683  
A:Experimental source: subspecies Queen Anne's Lace, isolate W001C  
C:Genetics:  
A:Gene: RAD23-1  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 49.9%; Score 954; DB 2; Length 382;  
Best Local Similarity 52.4%; Pred. No. 1.3e-56;  
Matches 205; Conservative 52; Mismatches 98; Indels 36; Gaps 9;

Qy 1 MKLTVKTLKGTFTFIRVQNPNDTMAVKNIEIQGKDSYPMGQQLIFNGKVLKDESTLE 60  
Db 1 MKIYVKTLGSKQFEIQVNPDDSVADVRSIETAGAAVYPAAGQMLIYQKVLKDGTTLL 60  
Qy 61 ENKVNEDGFLVWMLSKGKTSGSTSSQHSNTPATROAPPLRAPQAP-----QPVA- 114  
Db 61 ENNVAENSFVIMLSKSKSPSGSGSTTS-----TAAAPKAPQTSAPPSVPAPAVSQPPAST 116  
Qy 115 -PITTSQPEGLPAQAP-----NTHNAASNLLSGRNVDTIINQLMEMGGSGWDK 163  
Db 117 LPVPAPSPAPAPATAPIPSAAGSEANVYDSAAASLLVAGSLEGAIQIILDMGGGTWDRD 176  
Qy 164 KYQRAALAAVNNPERAVEVLYSGIPVTAETAVPI-----GGQGANTTDRAPTG--EAGLS 216  
Db 177 TVIRIVRAAFNNPERAVEVLYSGIPEQAE-APPVAPSPPSGQAANPLDQPPAAQAQAPAS 235  
Qy 217 GIPNAPLPLFPQAGSAG-AGAGGGPLDFLRNPNPQFQAVREVMHTNPQILQPLMVELSK 275  
Db 236 AGFNANPLDPLFPQGLPDMGSMAGAGNLDPLRTNQFQALRAMVQSNFQILQPLMLQELGK 295  
Qy 276 QNPQLRLIENHDFLQNLNPEPGEDEMPHAISSVTPEQEAIGRLSM 335  
Db 296 QNPHLMRLIQHQADFQLINPEMGEENLL-----GHGPOAISVTPEERDAIERLEAM 349  
Qy 336 GFDRAVIEAFLACDRNEELAANYLLEHAGE 366  
Db 350 GFDRELVLVEVFACNKEELAANYLLEHME 380

RESULT 4  
RAD23 protein homolog - rice  
C:Species: Oryza sativa (rice)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: T04150  
R:Schultz, T.F.; Quatrano, R.S.  
Plant Mol. Biol. 34, 557-562, 1997  
A:Title: Characterization and expression of a rice RAD23 gene.  
A:Reference number: 208695; MUID:97369378; PMID:9225866  
A:Accession: T04150  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-392 <SCH>  
A:Cross-references: EMBL:U63530; NID:g1488296; PIDN:AAB65841.1; PID:g1488297  
A:Experimental source: cv. Nipponbare  
C:Genetics:  
A:Gene: RAD23  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 46.5%; Score 887.5; DB 2; Length 392;  
Best Local Similarity 50.9%; Pred. No. 3.7e-52;  
Matches 199; Conservative 52; Mismatches 109; Indels 31; Gaps 10;

Qy 1 MKLTVKTLKGTFTFIRVQNPNDTMAVKNIEIQGKDSYPMGQQLIFNGKVLKDESTLE 60  
Db 1 MKISVTKLGSTFQIEVDQAQVADVVKRIETQGHYIPAEQMLIHQKVLKDDTTLD 60  
Qy 61 ENKVNEDGFLVWMLSKGKTSGSTSSQHSNTPATROAPPLRAPQAP-----QPVAPIIT 118  
Db 61 ENKVLNSFLVIMLRGKSGSSAPATSK-----APSNQAPPTQTVPAAPASQAPVAPATT 116





QY 1 MKLTVKTLKGTHTFEIRVQPNNDTIMAVKKNIEIEIQKDSYPMWQQLLIIFNGKVLKDBESTLE 60  
Db 1 MQVTLKTLQOQTKIDIDPBEITVAKLEKIESEKGDAPFVAGQKLIYAGKILNDDTALK 60  
QY 61 ENKVNEDGFLVLMVLSKGT-----SGSTGTSSSOHSNTPATQAP-----PLEAPQ 106  
Db 61 EYKIDENKFNVMVMTKPKAVSTPAPATTQOSAPASITAVTSSITTTTVAQAPTVPALAPT 120  
QY 107 QAP-----QP-PVAPITTSOPEGLPAQAP-----NTHDNA 135  
Db 121 STPASITPASATASSEPAPAAKQEPKPAETPVATSTATDTSOGSSRSNLFEDA 180  
QY 136 ASNLSSGRNVDTIINQLMWGGGSKDKVQRALRAAYNNPERAVEYLYSGIPVTABEIA 194  
Db 181 TSALVUTQGSYENVMVTEIMSG---YEREQVIAALRASFNPDRAVEYLLMGIPGDRSQ 237  
QY 195 VPIGGQGANNTDRAPTGAGLSGIPNTAPLDLFPQGSNAGGAGGGLDFLRNPNPQFA 254  
Db 238 VVDPPQAAT--GAPQSSAVAAAAATTT-----ATTTTSSGGHPLFLRNQPOFOQ 287  
QY 255 VREMVHTNPQILQPMVLVLSKONQIILBIENHDEFLQLINPEF-----GGEGDFLD 308  
Db 288 MRQIIQNPSSLPALLOQIGRENQLLOQISQHOEHFIQMLNEPVQAGQGGGGGSG 347  
QY 309 QPEDEMPHA--ISVTPPEQEAIGRLSMGDFRARVTEAFIACDRNEELAANYLLEHAGE 366  
Db 348 GIAEAGSGHMNYIQVTPQEKAEIRLKAIGPPEGLVITQYFACENENLAANFLQQNFD 407  
QY 367 ED 368  
Db 408 ED 409

## RESULT 8

S44443  
RAD23 protein homolog2 - human  
N:Alternate names: HRR23A protein  
C:Species: Homo sapiens (man)  
C:Date: 25-Dec-1994 #sequence\_revision 01-Sep-1995 #text\_change 18-Feb-2000  
C:Accession: S44443; T45076  
R:Maatani, C.; Sugawara, K.; Yanagisawa, J.; Sonoyama, T.; Ui, M.; Enomoto, T.; Takio, E.MBO J. 13, 1831-1843, 1994  
A:Title: Purification and cloning of a nucleotide excision repair complex involving the submitted to the EMBL Data Library, November 1996  
A:Reference number: S44345; MUID:94222030; PMID:8168482  
A:Accession: S44443  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-363 <MAS>  
A:Cross-references: EMBL:D21235; NID:g498145; PIDN:BAA04767.1; PID:d1005299; PID:g498148  
R:Lamerdin, J.; McCreedy, F.; Stilwogen, S.; Ramirez, M.; Carrano, A.  
A:Description: Characterization by genomic sequence analysis of a gene-rich 111 kb region submitted to the EMBL Data Library, November 1996  
A:Reference number: Z22906  
A:Accession: T45076  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-363 <LAM>  
A:Cross-references: EMBL:AD000092; PIDN:AAB51177.1  
A:Experimental source: cell line 5HL2-B; fibroblast  
C:Genetics:  
A:Gene: RAD23A  
A:Map position: 19p13.2  
A:Introns: 24/3; 78/3; 139/2; 158/1; 200/3; 227/1; 271/3; 326/3  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology  
F;3-82/Domain: ubiquitin homology <UBH>

Query Match 31.68; Score 603.5; DB 2; Length 363;  
Best Local Similarity 33.4%; Pred. No. 3.7e-33;  
Matches 133; Conservative 86; Mismatches 106; Indels 71; Gaps 8;  
QY 3 LTVKTLKGTHTFEIRVQPNNDTIMAVKKNIEIEIQKDSYPMWQQLLIIFNGKVLKDBESTLEEN 62  
Db 5 ITLKTLQOQTKIDIDPBEITVAKLEKIESEKGDAPFVAGQKLIYAGKILNDDVPIRDY 64

QY 63 KVNEDGFLVLMVLSKGTSGTSSSOHSNTPATQAPLEAPQAPQPPVAPIT----- 117  
Db 65 RIDENKFNVMVMT--KTRAGQGT-----APPEASPTAAPESSTSPAPTSMGSH 114  
QY 118 -----TSOPEGLPAQAPNT-----HDNAASNLSSGRNVDTIINQLMWEMG 156  
Db 115 PPAAREDKSPSESAPTTSPEVSQSVSSGSGREEDAASTLVGTSEYETMLTEIMSG 174  
QY 157 GGSWDRKDKVQRALRAAYNNPERAVEYLYSGIPVTABEIAVPIGGQGANNTDRAPTGAGLS 216  
Db 175 ---YERERVVAALRASYNPNPRAVEYLLTGIP-----GSPEPEHGSVQSSVS 219  
QY 217 GIPNTAPLDLFPQGSNAGGAGGGLDFLRNPNPQFAVREMVHTNPQILQPMVLVLSKQ 276  
Db 220 EQPATE-----AAGENPLEFLRDQPOQNMQRVITQONPALLPALLOQLQGE 285  
QY 277 NPQILRLBIENHDEFLQLINPE-----PEGEGDFLDQPEDEMPHAISVTPPEQEAIG 330  
Db 266 NPQLLOQISRHQQFIQMLNEPPGELADISDVEGVEGAIGEAPQMYIQVTPQEKAEIE 325  
QY 331 RLSEMGDFRARVTEAFIACDRNEELAANYLLEHAGEED 368  
Db 326 RLKALGPPESLVITQYFACENENLAANFLSQNFDE 363

## RESULT 9

T40115  
uv excision repair protein rad23 homolog - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 02-Sep-2000  
C:Accession: T40115; T51298  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Taylor, K.; Harris, D.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z21906  
A:Accession: T40115  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-368 <WOO>  
A:Cross-references: EMBL:AL031788; PIDN:CAA21170.1; GSPDB:GN00067; SPDB:SPBC2D10.12  
R:Zhao, Y.; Elder, R.T.  
submitted to the EMBL Data Library, July 1999  
A:Description: A fission yeast orthologue (rhp23) of the human nucleotide excision repair submitted to the EMBL Data Library, September 1998  
A:Reference number: Z25362  
A:Accession: T51298  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-368 <ZHA>  
A:Cross-references: EMBL:AF174293; PIDN:AAD51975.1  
A:Experimental source: strain SP223  
C:Genetics:  
A:Gene: rhp23; SPDB:SPBC2D10.12  
A:Map position: 2  
A:Introns: 23/3; 48/1; 328/3; 351/3  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 29.4%; Score 561.5; DB 2; Length 368;  
Best Local Similarity 36.0%; Pred. No. 2.4e-30;  
Matches 142; Conservative 70; Mismatches 127; Indels 55; Gaps 12;

QY 1 MKLTVKTLKGTHTFEIRVQPNNDTIMAVKKNIEIEIQKDSYPMWQQLLIIFNGKVLKDBESTLE 60  
Db 1 MNLTFKNLQOQKFVDSVADTKISELK--EKIQQNYEVEERQKLIYSGRILADDKTVG 58  
QY 61 ENKVNEDGFLVLMVLSKGTSGTSSSOHSNTPATQAPLEAPQAPQPPVAPIT--TT 118  
Db 59 EYNIKEQDFIVCMVSRPKTSTSTPKSAA----SPAPN--PPASVPBKVKVEASSTVAEST 112  
QY 119 SQPEGLPAQAPNTHDNAAGN-----LLSGRNVDTIINQLMWGGGSKDKVKQALR 170  
Db 113 STTQTVAAPSPNPDITATSEAPIDANTLAVGAQRNVAVENVMVG---YERSEVERAMR 169

```

171 AAYNNPERAVEYLYSGIP-----VTAEIAVPTGGCGANTTDRAPTGEAGLSGIPNTAPL 224
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
170 AAFNNPRAVEYLYLTGIPEDILNKRQREESAAALAAQQQSEALAPTSTG-----QPA 221
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
225 DLFPQGA-----SNAGCGGAGGGLDFLRNNPQFAVRWVHTNPQILPMLVELSKQN 277
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
222 NLFQQAALSENENQEPSNTVGDDPLGLFSITQFQQLRQIVQONPQMLSTIIQQIQGGD 281
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
278 PQILRLIENHDEFLQLLNPFEGEGDFLDQDPEDEMPHA---ISVTPPEQBAIGRLS 334
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
282 PALAQAITQNPFAFLQLL---AEGAEG-----ESALPSGGIQIQITQESSESIDLQ 331
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
335 MGFDRAVRVIAFLACDRNEELAAANYLLEHAGEED 368
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
332 LGFDNNVVIQAYLACDKNEELAAANYLFEHGEHE 365

RESULT 10
H86296
F309.1 protein - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C/Accession: H86296
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: H86296
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-113 <STO>
A/Cross-references: GB:AE005172; NID:g4966345; PIDN:AAD34676.1; GSPDB:GN00141
C/Genetics:
A/Map position: 1

```

[illegible]

RESULT 11  
T27774  
hypothetical protein ZK20.3 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T27774  
R:Gajadaty, S.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z20417  
A:Accession: T27774  
A:Status: preliminary; translated from GB/EMBL/DDBB  
A:Molecule type: DNA  
A:Residues: 1-372 <WIL>  
A:Cross-references: EMBL:Z69904; PIDN:CAA93780.1; GSPDB:GNO0020; CESP:ZK20.3  
A:Experimental source: clone ZK20  
C:Genetics:  
A:Gene: CESP:ZK20.3  
A:Map position: 2

A; Introns: 49/2; 73/3; 101/3; 369/3  
C; Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 21.2%; Score 405; DB 2; Length 372;  
Best Local Similarity 30.0%; Pred. No. 7.7e-20;  
Matches 111; Conservative 65; Mismatches 142; Indels 52; Gaps 8

QY 1 MKLIVTKLGTHFIRVQPNDTIMAVKKNIBEEIKDSDYPWGQQLLIFNGKVLKDESTLE 60  
:  
Db 52 LSVTFRTLTVNFNLNEDQTIAEVKALVASEKG-DDYAPELKLIYNGKILDDSVKVG 110  
:  
QY 61 ENKVNEDGFLVVMLSKGTSGTSTSSQHSNTPATQAAPLEAQQAPQPPVPAPTTSQ 120  
:  
Db 111 EVGFDSKSFVVVMLSKRKVTVEAPSSTVATAAEVPVAAPASNP--AAPAADVAPEAAA- 167  
:  
QY 121 PEGLPAAQNPTHNAANILLSGRNVDTIINQLMWGGGMDKKVKQVRALARAYNNPERAV 180  
:  
Db 168 ---PAAEAALTDOSEENVLA-----ITGMGYDRSQTIALARAFWNPDRAV 210  
:  
QY 181 EYLXGIPVTABIAVIPGGOGANTTDRAFTGEAGLSGITPTAPLDLPQGASNAGGGAGG 240  
:  
Db 211 EFLLNGLPDDA-----ADQEP-----DLGPEQINDVDEN 242  
:  
QY 241 GPLDFLRNPFQAVREWHVHTNPOLQPMVLVELSKNQPIURLRIEENHDEFLLQLNBPFE 300  
:  
Db 243 DDLNMLANMPOLAEIRALIQQNPMLAAVLQOLAANVPRLVQTIQNNQOAFMDDLINGCAQ 302  
:  
QY 301 GGEGDFLDOPED--EMPHATSVTPPEOEATGRLES--MGPDFRARVIEAFLACDRNEELAA 357  
:  
Db 303 GAGAAAGNAPERNTPRRHVIHLSPDEAAAIERIKAVVNAPEAVVVEAFACDKNSEAAI 362  
:  
QY 358 NYLLEHAGEE 367  
:  
Db 363 NFIFS NLDEE 372  
: :

RESULT 12  
S50507  
excision repair protein RAD23 - Yeast (Saccharomycetes cerevisiae)  
N; Alternate names: protein YEL037c  
C; Species: Saccharomycetes cerevisiae  
C; Date: 28-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 24-Sep-1999  
C; Accession: S50507; S30845; A54599; S38547  
R; Dietrich, F.S.  
submitted to the EMBL Data Library, December 1994  
A; Description: The sequence of S. cerevisiae cosmids 8199, 8334, and 9871.  
A; Reference number: S50491  
A; Accession: S50507  
A; Molecule type: DNA  
A; Residues: 1-398 <DI>  
A; Cross-references: EMBL:U18779; NID:g603625; PIDN:AAB65005.1; PID:g603642; MTR:  
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.;  
submitted to the EMBL Data Library, February 1993  
A; Reference number: S30812  
A; Accession: S30845  
A; Molecule type: DNA  
A; Residues: 1-398 <WL>  
A; Cross-references: GB:J25428; NID:g409246; PIDN:AAA16070.1; PID:g409247  
R; Watkins, J.F.; Sung, P.; Prakash, L.; Prakash, S.  
Mol. Cell. Biol. 13, 7757-7765, 1993  
A; Title: The Saccharomycetes cerevisiae DNA repair gene RAD23 encodes a nuclear  
A; Reference number: A54599; MUID:94067136; PMID:8246991  
A; Accession: A54599  
A; Molecule type: DNA  
A; Residues: 1-398 <WAT>  
A; Cross-references: GB:J25428; NID:g409246; PIDN:AAA16070.1; PID:g409247  
F; Melnick, L.; Sherman, F.  
J. Mol. Biol. 233, 372-388, 1993  
A; Title: The gene clusters ARC and COR on chromosomes 5 and 10, respectively,  
A; Reference number: S38543; MUID:94016558; PMID:8411151  
A; Accession: S38547  
A; Molecule type: DNA  
A; Residues: 1-276, 'R', 278-398 <MEL>

A, A: Cross-references: EMBL:566117; NID:G430823; PIDN:ARB28441.1; PID:G430823  
C: Genomics:  
A: Gene: SGD:RAD23  
A: Cross-references: SGD:S0000763; MIPS:YEL037C  
A: Map position: 5L  
C: Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology  
C: Keywords: DNA repair  
F: 2-77/Domain: ubiquitin homology <UBH1>

Query Match	21.1%	Score 403;	DB 2;	Length 398;
Best Local Similarity	31.7%	Pred. NO. 1.1e-19;		
Matches 130;	Conservative	63;	Mismatches 157;	Indels 60;
Gaps				16;

Qy	1	MKLTVKTLKGT	FEIR	QPN	DTI	MAV	-	KNT	EE	IO	G	K	D	S	Y	P	W	G	O	L	L	I	F	N	G	K	L	D	E	S	T	L	59																					
Db	2	VSLTFNFKK	KEK	VP	LD	LE	P	S	T	I	L	E	T	T	K	L	A	Q	S	I	C	E	S	E	---	Q	I	K	L	I	S	K	V	Q	D	S	T	57																
Qy	60	EENKVN	D	E	G	F	L	V	M	L	S	K	G	T	S	G	T	S	S	O	R	S	N	T	P	A	T	R	Q	A	P	P	L	E	A	P	O	A	---	P	O	P	P	V	A	P	I	116						
Db	58	SECG	L	K	G	D	Q	V	V	F	M	S	Q	K	---	S	I	K	K	T	E	P	P	I	A	P	E	S	A	T	T	P	C	R	N	S	T	E	A	S	P	D	A	A	P	A	114							
Qy	117	TT	---	S	O	P	-	E	G	L	P	A	Q	A	P	N	T	H	N	A	S	N	L	L	S	R	N	V	D	T	I	I	N	Q	L	M	E	M	G	G	S	D	K	D	K	V	O	R	A	L	A	171		
Db	115	ATA	P	E	G	S	O	P	Q	E	O	A	T	T	E	R	T	E	S	A	S	T	P	G	F	V	G	T	E	R	N	E	T	I	E	M	E	M	G	---	Y	O	R	E	E	V	E	R	A	L	A	171		
Qy	172	AYNN	P	R	A	V	E	L	Y	G	S	I	P	T	A	I	A	I	P	I	G	O	G	A	N	T	T	D	R	A	P	T	G	E	A	G	L	S	G	I	P	N	T	A	P	L	D	I	F	P	O	A	231	
Db	172	AFNN	P	R	A	V	E	L	L	M	G	I	P	-	E	N	L	R	O	P	E	Q	O	A	A	A	E	Q	P	S	T	A	T	T	A	E	Q	P	-	A	E	D	D	I	F	A	Q	A	227					
Qy	232	--	SN	A	G	G	A	---	G	G	P	L	D	F	L	R	N	P	---	Q	F	A	V	R	E	M	V	H	T	T	P	Q	I	L	O	P	M	L	S	K	Q	N	277											
Db	228	O	G	G	N	A	S	G	A	L	T	T	G	A	T	D	A	A	O	G	P	P	G	S	T	G	L	T	V	E	D	L	L	S	L	Q	V	S	G	N	E	A	L	A	P	L	E	N	I	S	A	287		
Qy	278	Q	I	L	R	L	I	E	N	H	E	F	L	L	N	E	P	---	F	E	G	---	E	G	D	F	L	---	Q	P	E	D	E	M	315																			
Db	288	P	O	L	R	H	I	M	A	N	E	V	F	S	M	L	L	E	A	V	C	N	D	M	Q	D	M	E	G	A	D	D	W	E	G	E	I	E	V	T	E	C	A	A	A	G	L	G	O	G	E	C	B	347
Qy	316	P	H	A	I	S	V	T	P	E	O	E	A	T	G	R	L	S	M	G	F	D	R	A	V	I	E	A	F	L	A	C	R	N	E	E	L	A	N	Y	L	L	-	E	H	A	364							
Db	348	S	F	O	V	D	T	T	P	E	D	O	A	T	S	R	L	C	E	L	G	E	P	R	D	L	T	Q	V	F	A	C	D	K	N	E	E	A	A	N	I	L	F	S	D	H	A	397						

```

RESULT 13
T51479
Hypothetical protein T21H19_10 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51479
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata
submitted to the Protein Sequence Database, August 2000
A:Reference number: Z25394
A:Accession: T51479
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-142 <SAT>
A:Cross-references: EMBL:AL391148
A:Experimental source: cultivar Columbia, BAC clone T21H19
C:Genetics:
A:Map position: 5
A:Introns: 22/3; 76/3; 111/2
A>Note: T21H19_10

Query Match      10.7%; Score 204.5; DB 2; Length 142;
Best Local Similarity 33.8%; Pred. No. 5.8e-07;
Matches 51; Conservative 22; Mismatches 54; Indels 25; Gaps 3

QY    1 MKLVTKLTGTFEIRVQPNDITMAVKKNIEETGGKDSYPWGQQLLIFNGKVLKDESTLE 60
      ||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| : ||||| :
Db     1 MKLIIVTLKGIRFEIEVKPNDSVAEVKKNIETVGWAGSEYPAQAQILLIHKEKLRLDETME 60

QY    61 ENKVNEDGFLVWLMSKGKTSGSTGTSSSQHSNTPATRAQPPLEAPQAQPPVPAPITT-- 118
      ||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db     61 ANKVFQKSVALIITIGCLE-----EMEKQNPLFQMIRHNAGFPVPLNKE 160

```

```

Qy      119  -----SQPEG--LPAQAPNTHDNASNLL 140
          :||| |   |   |   |   |   |   |
Db      108  SFRDNELAQPEEDLLQLQTVAVDEAINRVL 139

RESULT 14
C84549
probable ubiquitin-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84549
M:Llin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii
R.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; T
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.;
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84549
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-551 <STO>
A:Cross-references: GB:AE002093; NID:g4584343; PIDN:AAD25138.1; GSPDB:GNO0139
C:Genetics:
A:Map position: 2

```

	Query Match	10.5%	Score 201;	DB 2;	Length 551;
	Best Local Similarity	19.8%	Pred. No. 6.2e-06;		
	Matches 112;	Conservative 62;	Mismatches 157;	Indels 234;	Gaps 21
QY	3	LTVKTLKGTHTFEIRVQPNDDTIMAVKKNTEETIGQKDSYPMWGQOLIFNGKVLKDESTLEEN	62		
DB	20	VNIRCSNGTKFSVTKSLDSTVESFK---ELVAQSSDVPAHQORLIYKGRILKDDQTL---	73		
QY	63	KVNEDGFLVWLMSKGTSGST---GTSSSQHSNTPATROAPLEAPQAPQPPV---	113		
DB	74	-----LSYGLQADHTHTWYRGSAPSAP-----PPAPASQTTAPSVTRGVGSD	117		
QY	114	-----APITTSQP-----EGLP-----AQAPN-----	130		
DB	118	NSSNIGGASPGESLSPGLGCFNPLGGNAMSGLFGAGLPDLVOTOOLAAQNPNHIRMNT	177		

```

131  -THDNASH-----LLSGRVDTIINQLEMGGGWDKQKVQRAALRAAYN----- 178
178  PAIQNLMMNPFWRSIMMNPQMRELVDNRNPELGHVLDNPSTILQRTLEAARNPELMREMM 237
175  -NPERA-----VEYLSGIPVTAIEAIPIGG----- 199
238  RNTDRAMSIENMPEGFNMLRMRYENVQPELNNATTMSGNNTGSPNPFALLGNQGV 297
200  -QGANTTDRAPTGEAGLSGIPNTAPL-----DLFPQGASNAG----- 235
298  TQGSDDASNNSTPNAGTGTIPNANPLPNWGATGGOTTAPGRNTVGGDARSPLGGLGGL 357
236  ---GGAGG-----GPLDFLRNNPQAVREVMHTNPQILOPMLVLSKQ 276
358  GSLGGGLGGLMGADSPLGATPDASQLSQLLQNPASISOMQSVFSNPQYMN-----QLMSL 413
277  NPQILBLIENHDEFQLLNEPPEGEGDFLDQPEDEMPHAI----- 319
414  NPQLRMLDSN-PQLRENNQNP-----DFLRQFSSPEMMQQMMTTLQQLSQNRNTASQD 466
320  -----SVTPEEQEA--IGRLSMGP-DRAR 341
467  AGQTGAATGNNGGLDLLMMFMSGLAGGLSGTGNQSNVPPEERYATQLQLOEQMGFYDRAE 526
342  VIEAFLACDRNEELAANYLLEHAGE 366
527  NIRALLATNGNVNAAVERLLGSIGQ 551

```

## RESULT 15

Search completed: December 17, 2003, 06:20:49  
Job time : 17.6624 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:16:39 ; Search time 10.4735 Seconds  
(without alignments)  
1652.347 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 1910

Sequence: 1 MKLTVTKLKGTHFEIRVQPN.....CDRNEELAANYLLEHAGEED 368

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	634	33.2	416	1 R23B_MOUSE	P54728 mys musculus
2	625.5	32.7	409	1 R23A_HUMAN	P54727 homo sapien
3	603.5	31.6	363	1 R23B_MOUSE	P54725 homo sapien
4	597.5	31.3	363	1 R23A_MOUSE	P54726 mys musculus
5	561.5	29.4	368	1 RH23_SCHPO	O74803 schizosacch
6	403	21.1	398	1 RA23_YEAST	P32628 saccharomyc
7	163.5	8.6	354	1 YAUG_SCHPO	Q10169 schizosacch
8	154	8.1	373	1 DSK2_YEAST	P48510 saccharomyc
9	122.5	6.4	76	1 UBIQ_DROME	Q9V214 drosophila
10	122.5	6.4	76	1 UBIQ_HUMAN	P02248 homo sapien
11	122	6.4	979	1 RFX1_HUMAN	P22670 homo sapien
12	121.5	6.4	81	1 NED8_HUMAN	Q15843 homo sapien
13	121.5	6.4	81	1 NED8_MOUSE	P29595 mys musculus
14	120.5	6.3	76	1 UBIQ_EIWEHO	P46574 eimeria bov
15	118.5	6.2	76	1 UBIQ_LEIMA	Q05550 leishmania
16	118.5	6.2	76	1 UBIQ_PHYIN	P22589 phytophthor
17	118.5	6.2	76	1 UBIQ_STRPU	P23398 strongyloce
18	117.5	6.2	76	1 UBIQ_LEITA	P49635 leishmania
19	116.5	6.1	76	1 UBIQ_ACACA	P49634 acanthamoeb
20	116.5	6.1	76	1 UBIQ_NEUCR	P31117 neurospora
21	115.5	6.0	93	1 UBIQ_NPVOP	Q05120 argyria pseu
22	114.5	6.0	76	1 UBIQ_AGLNE	P42740 aglaetahamni
23	114.5	6.0	76	1 UBIQ_CABEL	P14792 caenorhabdi
24	114.5	6.0	76	1 UBIQ_GEOCY	P59669 geodia cydo
25	114.5	6.0	76	1 UBIQ_YEAST	P04838 saccharomyc
26	113.5	5.9	76	1 UBIQ_ACACL	P42739 acetabulari
27	113.5	5.9	76	1 UBIQ_CHLRE	P14624 chlamydomon
28	113.5	5.9	76	1 UBIQ_COPCO	P19848 coprinus co
29	113.5	5.9	76	1 UBIQ_TRYCR	P08565 trypanosoma
30	113.5	5.9	78	1 UBL1_SCHPO	Q14399 schizosacch
31	113	5.9	392	1 RHE1_HUMAN	Q05925 homo sapien
32	113	5.9	989	1 FRQ_NEUCR	P19970 neurospora
33	112.5	5.9	76	1 UBIQ_ARATH	P59263 arabidopsis

## ALIGNMENTS

### RESULT 1

ID	R23B_MOUSE	STANDARD;	PRT;	416 AA.
AC	P54728;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	UV excision repair protein RAD23 homolog B (MHR23B) (XP-C repair complementing complex 58 kDa protein) (P58).			
GN	RAD23B OR MHR23B.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RP	SEQUENCE FROM N.A.			
RN	STRAIN=BALB/c; TISSUE=Testis;			
RC	MEDLINE=96403997; PubMed=8808275;			
RX	van der Spek P.J., Visser C.E., Hanaoka F., Smit B., Hagemeijer A., Bootsma D., Hoeijmakers J.H.J.;			
RA	"Cloning, comparative mapping, and RNA expression of the mouse homologues of the Saccharomyces cerevisiae nucleotide excision repair gene RAD23.";			
RT	Genomics 31:20-27(1996).			
RL	-/- FUNCTION: INVOLVED IN DNA EXCISION REPAIR. MAY PLAY A PART IN DNA DAMAGE RECOGNITION AND/OR IN ALTERING CHROMATIN STRUCTURE TO ALLOW ACCESS BY DAMAGE-PROCESSING ENZYMES.			
CC	-/- SUBUNIT: HETERODIMER OF A 125 kDa SUBUNIT (P125) AND OF A 58 kDa SUBUNIT (P58).			
CC	-/- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	-/- SIMILARITY: Contains 1 ubiquitin-like domain.			
CC	-/- SIMILARITY: Contains 2 UBA domains.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; X92411; CAA63146.1; -.			
DR	HSSP; P54725; IDV0.			
DR	MGD; MGI:105128; Rad23b.			
DR	InterPro; IPR004806; Rad23.			
DR	InterPro; IPR006636; STI1.			
DR	InterPro; IPR000449; UBA_domain.			
DR	InterPro; IPR000626; Ubiquitin.			
DR	Pfam; PF00627; UBA; 2.			
DR	Pfam; PF00240; ubiquitin; 1.			
DR	SMART; SM00727; STI1; 1.			
DR	SMART; SM00165; UBA; 2.			
DR	SMART; SM00213; UBQ; 1.			
DR	TIGRFAMs; TIGR00601; rad23; 1.			
DR	PROSITE; PS0030; UBA; 2.			
DR	PROSITE; PS00053; UBIQUITIN_2; 1.			

P03993 glycine max  
P20685 tetrahymena  
P46379 homo sapien  
P09853 herpes simp  
P23324 euplates eu  
P15174 trypanosoma  
P08618 dictyosteli  
P14283 bordetella  
P77965 synechocyst  
P48681 homo sapien  
P45118 haemophilus  
Q13495 homo sapien

34 112.5 5.9 76 1 UBIQ\_SOYBN  
35 112.5 5.9 76 1 UBIQ\_TETPY  
36 111 5.8 1132 1 BAT3\_HUMAN  
37 111 5.8 1144 1 RIR1\_HSV23  
38 110.5 5.8 76 1 UBIQ\_EUPEU  
39 109.5 5.7 76 1 UBIQ\_TRYBB  
40 108.5 5.7 76 1 UBIQ\_DICDI  
41 108 5.7 910 1 PERT\_BORPE  
42 108 5.7 1102 1 RPOB\_SYNY3  
43 107.5 5.6 1618 1 NEST\_HUMAN  
44 107 5.6 567 1 ODP2\_HAEIN  
45 106.5 5.6 701 1 CGI\_HUMAN

```
KW DNA damage; DNA repair; Nuclear protein; Repeat.
FT DOMAIN 1 79 UBIQUITIN-LIKE.
FT DOMAIN 188 228 UBA 1.
FT DOMAIN 371 411 UBA 2.
FT DOMAIN 255 261 POLY-ALA.
FT DOMAIN 262 270 POLY-THR.
FT DOMAIN 336 355 POLY-GLY.
SQ SEQUENCE 416 AA; 43516 MW; 1350245AGD892205 CRC64;

Query Match 33.2%; Score 634; DB 1; Length 416;
Best Local Similarity 33.5%; Pred. No. 3.2e-33;
Matches 146; Conservative 93; Mismatches 109; Indels 88; Gaps 11;

Qy 1 MKLVTKLKGTHFIRVOPNDTIMAVKKNIEIIGKDSYPMGQQLLIPNGKVLKDESTLE 60
Db 1 MQVLTKLQQQTFKIDIDPEETVKALKEIESEKGDAPFVAGQKLIYAGKILSDTALK 60

Qy 61 ENKNEDEFLVMSKGTSGTSSQHNTPA---TROAPPLEAPQ-CARPPVPAPI 116
Db 61 EYKIDENFVVVMTKPAVTTAVDATTQPSSTPSTTVSSSPAVAAQAAPATPALAPT 120

Qy 117 TT-----SQPE---GLPAQAP-----NTHDNA 135
Db 121 STPASTTASTASSEPAAGATQEKPAEKPAQTPLVTSAPADSTPCDSRSNLPEDA 180

Qy 136 ASNLISGRNVDTIINQLMEMGGGWDKQVORALRAAYNNPERAVEYLYSGIPVTAETAV 195
Db 181 TSALVTGQSYENVMTIEISMVG---YERQVIAALRASENNPDRAVEYLLMGIPGDRSQ 237

Qy 196 PTGGGQANTDRAPTGERAGLGIPTNAPLIDLPQQAASVAG-----CGAGGGLDPLRN 248
Db 238 VVD-----PPQAVSTGTPQS-----PAVAAAAATTTATTTTSGHPLFLRN 281

Qy 249 NPQFQAVRMVHTNPQILQPLMVELSKNPQILRIEENHDFLOLLNEPFE----- 300
Db 282 QPQFQOQMOITIQNPSLLPALLQQLGRENPLQQLISQHQEFIQWLNEPVEAGQGQGG 341

Qy 301 -----GGEGDFLDQDEEMPHAISVTPEQEAIGRESMGFDRARVIEAFIACDRN 352
Db 342 GGGGGGGGGGGGGAEGAGSGHNY-IQVTPQKEAIERLKALGPPEGLVIQAYFACEKN 400

Qy 353 BELAANYLLEHAGEED 368
Db 401 ENLAANFLQQNFED 416

RESULT 2
R23B_HUMAN STANDARD; PRT; 409 AA.
AC P54727; Q8WUB0;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UV excision repair protein RAD23 homolog B (HHR23B) (XP-C repair
DE complementing complex 58 kDa protein) (p58).
DN RAD23B.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94222030; PubMed=8168482;
RA Masutani C., Sugawara K., Yanagisawa J., Sonoyama T., Ui M.,
RA Enomoto T., Takio K., Tanaka K., van der Spek P.J., Bootsma D.,
RA Hoeijmakers J.H.J., Hanaoka F.;
RT "Purification and cloning of a nucleotide excision repair complex
RT involving the xeroderma pigmentosum group C protein and a human
RT homologue of yeast RAD23";
RL EMBO J. 13:1831-1843(1994).
[2]
RN SEQUENCE FROM N.A., AND VARIANT VAL-249.
RP Rieder M.J., Livingston R.J., Daniels M.R., Montoya M.A., Chung M.-W.,
RA
```

```
RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Wittrak L.A., Nickerson D.A.;
RA Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RA Ramsay H.;
RN Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A., AND VARIANT VAL-249.
RC TISSUE=Uterus;
RX MEDLINE=2388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haiech F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length
RA human and mouse cDNA sequences";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[1] FUNCTION: INVOLVED IN DNA EXCISION REPAIR. MAY PLAY A PART IN DNA
CC DAMAGE RECOGNITION AND/OR IN ALTERING CHROMATIN STRUCTURE TO
CC ALLOW ACCESS BY DAMAGE-PROCESSING ENZYMES.
CC [2] SUBUNIT: Heterodimer of a 125 kDa subunit (p125) and of a
CC 58 kDa subunit (p58). Interacts with MJD.
CC [3] SUBCELLULAR LOCATION: Nuclear (Probable).
CC [4] DOMAIN: The ubiquitin-like domain mediates interaction with MJD.
CC [5] SIMILARITY: Contains 1 ubiquitin-like domain.
CC [6] SIMILARITY: Contains 2 UBA domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D21090; BAA04652.1; -
CC EMBL; AL137852; CAD13275.1; -
CC EMBL; AY165178; AAN47194.1; -
CC EMBL; BC020973; AAH20973.1; -
CC FIR; S44346; S44346.
CC HSSP; P54725; 1DVO.
CC Genew; HGNC:9813; RAD23B.
CC MIM; 600062; -
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0003697; F:single-stranded DNA binding activity; TAS.
CC GO; GO:0006289; P:nucleotide-excision repair; TAS.
CC InterPro; IPR004806; Rad23.
CC InterPro; IPR006636; STI1.
CC InterPro; IPR000449; UBA domain.
CC InterPro; IPR000626; Ubiquitin.
CC Pfam; PF00627; UBA; 2.
CC Pfam; PF00240; ubiquitin; 1.
CC SMART; SM00727; STI1; 1.
CC SMART; SM00165; UBA; 2.
CC SMART; SM00213; UHQ; 1.
CC TIGRFAMs; TIGR00601; rad23; 1.
CC PROSITE; PS50030; UBA; 2.
CC PROSITE; PS50053; UBIQUITIN 2; 1.
CC DNA damage; DNA repair; Nuclear protein; Polymorphism; Repeat.
KW
```





```

DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; UbiQuitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00727; STIL; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS50030; UBA; 2.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
KW DNA damage; DNA repair; Nuclear protein; 3D-structure; Repeat.
FT DOMAIN 1 81 UBIQUITIN-LIKE.
FT DOMAIN 161 201 UBA 1.
FT DOMAIN 318 358 UBA 2.
FT HELIX 321 324
FT TURN 325 331
FT HELIX 334 341
FT TURN 342 345
FT HELIX 348 354
FT TURN 355
SQ SEQUENCE 363 AA; 39609 MW; C4E47B9313BB47B5 CRC64;

Query Match 31.6%; Score 603.5; DB 1; Length 363;
Best Local Similarity 33.4%; Pred. No. 2,3e-31;
Matches 133; Conservative 88; Mismatches 106; Indels 71; Gaps 8;

QY 3 LTVKTLKGTHEIRVQNDTIMAVKNIEEIQKDSYPMGQQLIFNGKVLKDESTLEEN 62
DB 5 ITLKTQQQTKIRMEPDETVMKLEKEAEKGRDAFPVAGQKLIYAGKILSDVPIRDY 64

QY 63 KVNEDGFLVMSLKGKTSSTSSQHSNTPATRAQAPLEAPQAPQAPVAPIT----- 117
DB 65 RIDENKFNVMVMTKAK-----APPEASPTAPESSTSPFPAPTSGMSHP 114

QY 118 -----TSQPEGLPAQAPNT-----HDNAASNLGSRNVDITINOLMWMG 156
DB 115 PPAAREDKSPSEAPPTSPESVSGSSGSGREDAASTLVGSEVETMLTMSMG 174

QY 157 GSGWDKQVQALRAAYNNPRAVEYLXGIPVTABIAVPIGGQAGTDDRAPTGAGLS 216
DB 175 -----YERERVVAALRASYNPPRAVEYLLTGIP-----GSPEHGSVQESQVS 219

QY 217 GIPNTAPLDFPQASNAGGAGGGLDFLRNNPQFQAVREVMVHTNPQILQPMVLKSLQ 276
DB 220 EQPATE-----AAGENPLEFLRDPQFQNNRQVIQQNPALLPALLOQLQGE 265

QY 277 NQILRLIEENHDEFLQLNEP-----PEGEGDFLDQPEDEMPHAISVTPPEQATG 330
DB 266 NPQLQQLISRHQEQFQMLNEPPEGLADISDVEGEVGAIGEAPQMNVIQVTPQKEALE 325

QY 331 RLESMDGFRARVIEAFACDRNEELAANYLLEHAGEED 368
DB 326 RLKALGFPPESLVIQAYFACEKKNENLAANFLLSQNFDE 363

RESULT 4
R23A_MOUSE
ID R23A_MOUSE STANDARD; PRT; 363 AA.
AC P54726;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UV excision repair protein RAD23 homolog A (MHR23A).
GN RAD23A OR MHR23A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Testis;
RX MEDLINE=96403997; PubMed=8808275;
RA van der Spek P.J., Visser C.B., Hanaoka F., Smit B.,

```

```

RA Hagemeijer A., Bootsma D., Hoeijmakers J.H.J.;
RT "Cloning, comparative mapping, and RNA expression of the mouse
RT homologues of the Saccharomyces cerevisiae nucleotide excision repair
RL Gene RAD23.";
CC Genomics 31:20-27(1996).
CC -|- FUNCTION: Involved in postreplication repair of UV-damaged DNA.
CC Postreplication repair functions in gap-filling of a daughter
CC strand on replication of damaged DNA (Potential).
CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -|- SIMILARITY: Contains 1 ubiquitin-like domain.
CC -|- SIMILARITY: Contains 2 UBA domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X92410; CAA63145.1; -.
DR HSSP; P54725; LDV0.
DR MGD; MGI:105126; Rad23a.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STIL.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; UbiQuitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00727; STIL; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS50030; UBA; 2.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
KW DNA damage; DNA repair; Nuclear protein; Repeat.
FT DOMAIN 1 79 UBIQUITIN-LIKE.
FT DOMAIN 161 201 UBA 1.
FT DOMAIN 318 358 UBA 2.
SQ SEQUENCE 363 AA; 39769 MW; 67EAB96EBBASC203 CRC64;

Query Match 31.3%; Score 597.5; DB 1; Length 363;
Best Local Similarity 33.7%; Pred. No. 5,5e-31;
Matches 136; Conservative 84; Mismatches 101; Indels 83; Gaps 11;

QY 3 LTVKTLKGTHEIRVQNDTIMAVKNIEEIQKDSYPMGQQLIFNGKVLKDESTLEEN 62
DB 5 ITLKTQQQTKIRMEPDETVMKLEKEAEKGRDAFPVAGQKLIYAGKILSDVPIRDY 64

QY 63 KVNEDGFLVMSLKGKTSSTSSQHSNTPATRAQAPLEAPQAPQAPVAPIT----- 116
DB 65 HIDEKFNVMVMTKAK-----APPEASPTAPESSTSPFPVPLASG 110

QY 117 -----TTQPEGLPAQAPNT-----HDNAASNLGSRNVDITINOL 152
DB 111 MSHPPPTSRDKSPSESTTTTSPESVSGSSGSGREDAASTLVGSEVETMLTETI 170

QY 153 MEMGGSDKQVQALRAAYNNPRAVEYLXGIPVTABIAVPIGGQAGTDDRA--PT 210
DB 171 MSMG---YERERVVAALRASYNPPRAVEYLLTGIPGSP-----PEHG-SVQESQRAEQPA 223

QY 211 GEAGLSGIPNTAPLDFPQASNAGGAGGGLDFLRNNPQFQAVREVMVHTNPQILQPM 270
DB 224 TEA-----AAGENPLEFLRDPQFQNNRQVIQQNPALLPAL 259

QY 271 VELSKNPQILRLIEENHDEFLQLNEP-----PEGEGDFLDQPEDEMPHAISVTP 324
DB 260 QQLQGENPQLLQQLISRHQEQFQMLNEPPEGLADISDVEGEVGAIGEAPQMNVIQVTP 319

QY 325 EQEATGRLESMDGFRARVIEAFACDRNEELAANYLLEHAGEED 368
DB 320 EKEATERLKALGFPPESLVIQAYFACEKKNENLAANFLLSQNFDE 363

```



modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

EMBL; AF174293; AADS1975.1; -  
 EMBL; AL031788; CAA21170.1; -  
 PIR; T40115; T40115.  
 KSSP; P54725; IDV0.  
 GenDB SPombe; SPBC2D10.12; -  
 InterPro; IPR004806; Rad23.  
 InterPro; IPR006636; STI1.  
 InterPro; IPR000449; UBA domain.  
 InterPro; IPR000626; Ubiquitin.  
 Pfam; PF00627; UBA; 2.  
 Pfam; PF00240; ubiquitin; 1.  
 SMART; SM00727; STI1; 1.  
 SMART; SM00165; UBA; 2.  
 SMART; SM00213; UBQ; 1.  
 TIGR PAM; TIGR00601; rad23; 1.  
 PROSITE; PS50030; UBA; 2.  
 PROSITE; PS50053; UBIQUITIN 2; 1.  
 DNA damage; DNA repair; Nuclear protein; Repeat.  
 DOMAIN 1 77 UBIQUITIN-LIKE.  
 DOMAIN 135 185 UBA 1.  
 DOMAIN 320 360 UBA 2.  
 DOMAIN 119 122 POLY-ALA.  
 DOMAIN 205 208 POLY-GLN.  
 SEQUENCE 368 AA; 40135 MW; 5CE75EB7E190EFD4 CRC64;

Query Match 29.4%; Score 561.5; DB 1; Length 368;  
 Best Local Similarity 36.0%; Pred. No. 1.1e-28;  
 Matches 142; Conservative 70; Mismatches 127; Indels 55; Gaps 12;

QY 1 MKLTVKTLKGFHFEIRVPQNDITNAVKNTEIEIQKDSYPMGQQLIFNGKVLKDESTLE 60  
 DB 1 MLLTFNQLQQKFKVISDSVADTKISELK-EKIQQNYEVEKQKLYSGRIADDKTVG 58

QY 61 ENKVNEDGLVVLMSKGTSGTSSQHSNTPATRAQAPPEAPQAPQAPPPVAPI--TT 118  
 DB 59 EYNIKEQDFIVCMVSRPKTSTSTPKSAA----SPAPN--PPASVPEKKVEAPSTVAEST 112

QY 119 SQPEGLPQAQANTHNDNAASN-----LLSGRNVDTIINQLMVGSGGSDKKVQRALR 170  
 DB 113 STQTQVAANAAPSNFDTTTSBAPIDANTLAVGAQRNVAVENVMEMG---YERSEVERAMR 169

QY 171 AAYNNPRAVEVLYSGIP-----VTAEIAPVIGGQANTTTDRAPTGEAGLSGIPNTAPL 224  
 DB 170 AAFNNPRAVEVLYTGIPEDILNQRSESAALAAQQQSEALAPTSTG-----QPA 221

QY 225 DLFFPGA-----SNAGGAGGGLDFLRNNPQFQAVRMVHTNPOILQPMVLVELSKQN 277  
 DB 222 NLFEQAALSENENQEQPSNTVGGDDPLGLFLRSIPQFQLRQIVQONFQMLTILQIQGQD 281

QY 278 POILRLIENHDEFLQLNPEFEGEGDFLDQDPEDEMPHA---ISVTPEQRAIGRLS 334  
 DB 282 PALAQAITQNPEAFQLLL---AEGAEG-----ESALPSGGIQLIQITQSESESIDRLCQ 331

QY 335 MGFDRAVIEAFACDRNEELAANYLLEHAGEED 368  
 DB 332 LGFDRNVIQIAYLACDKNEELAANYLFEHGESE 365

RESULT 6  
 RA23\_YEAST  
 ID RA23\_YEAST STANDARD; PRT; 398 AA.  
 AC P32628;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE UV excision repair protein Rad23.  
 GN RAD23 OR YEL037C OR SYGP-ORP29.  
 OC Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;



RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of *Schizosaccharomyces pombe*."  
 RL Nature 415:871-880 (2002).  
 CC -1- SIMILARITY: TO YEAST DSK2.  
 CC -1- SIMILARITY: Contains 1 ubiquitin-like domain.  
 CC -1- SIMILARITY: Contains 1 UBA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z69240; CAA93239.1; --  
 DR PIR: T38404; T38404.  
 DR GenBank: SPombe; SPAC26A3.16; --  
 DR InterPro: IPR006636; ST11.  
 DR InterPro: IPR000449; UBA domain.  
 DR InterPro: IPR006626; Ubiquitin.  
 DR Pfam: PF00627; UBA; 1.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR SMART: SM00727; ST11; 2.  
 DR SMART: SM00165; UBA; 1.  
 DR SMART: SM00213; UBO; 1.  
 DR PROSITE: PS50030; UBA; 1.  
 DR PROSITE: PS50053; UBIQUITIN\_2; 1.  
 KW Hypothetical protein.  
 FT DOMAIN 1 78 UBIQUITIN-LIKE.  
 FT DOMAIN 309 353 UBA.  
 SQ SEQUENCE 354 AA; 36819 MW; 1A99B2D97E73A831 CRC64;

Query Match 8.6%; Score 163.5; DB 1; Length 354;  
 Best Local Similarity 22.7%; Pred. No. 0.0016;  
 Matches 95; Conservative 62; Mismatches 134; Indels 127; Gaps 21;  
 QY 1 MKLTVKLTAGTHFEIRVQPNDDTMAVKNIIEIQKDSYFPGQQLLIIFNGKVLKDESTLE 60  
 DB 4 ISLTIAANDQKYAVTVDSESSVLLALKEAIPVADIEK---ERQRLIYAGRVLKDEESLK 60  
 QY 61 ENKVNEDGLVLMSSKGKTSKTSSTSSQHSNTPTATROAPLEAPQAPQPPVAPITTSQ 120  
 DB 61 TYKIQDGHSHILVKLTIGQNPAAATVSDR-----TQVPTNIQAGCGANFPLNLTAR 115  
 QY 121 PEG----LPAQA-----PNTHDNAASNLISGRNVDTIIN-----QLMEW---- 155  
 DB 116 YAGFNIPMPSASMFQGNPNPVPSTTELA--NMLSNPMVQSSINEMFSNPQMLDMLINS 173  
 QY 156 -----GGGSWDKQVQR-ALRAAYNPE--RAVEYLY-----SGIPVTABIAVPIGGQGA 202  
 DB 174 SPHLRNAPPYVVRQMOSPEFRAMTDPDPTMRQMAQLHQMGGAAGIDPMSLMGGGLGAGL 233  
 QY 203 NTTDRAPTCEAGLSGIPNTAPLDLPQGSNAGGG-AGGGLDPLRNPNPQFAVREWHVT 261  
 DB 234 -----GGLGAGLGEP-----GGANNATAGIAGAAPVD-----QTAAANTI--- 269  
 QY 262 NPQILQPLMLVELSKNPQLRLIEENHDFLOLNE-----PPEGGEED-----F 306  
 DB 270 -----QN-----LLNLLGAGFGAGLGDAGLGGAGGAS 299  
 QY 307 LQPEDENPHAIISTVPEIQ--EATGRLESMTGF-DRARVIEAFLACDRNEELAAANYLL 361  
 DB 300 PPAPAQDTRP-----PEERYAEQLSQLNEMGFVDFERNVQALRRSGGNVQAIESLL 351

## RESULT 8

DSK2\_YEAST  
 ID DSK2\_YEAST STANDARD; PRT; 373 AA.  
 AC P48510;  
 DT 01-FEB-1996 (Rel. 33, Created)

Query Match 8.1%; Score 154; DB 1; Length 373;  
 Best Local Similarity 23.2%; Pred. No. 0.0067;  
 Matches 84; Conservative 50; Mismatches 142; Indels 86; Gaps 18;

QY 1 MKLTVKLTAGTHFEIRVQPNDDTMAVKNIIEIQKDSYFPGQQLLIIFNGKVLKDESTLE 59  
 DB 1 MSLNIHKSQDKWEVNVVAPSTVLQFKEAKANG---IPVANORLIYSGLKLDQTV 57

DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Ubiquitin-like protein DSK2.  
 GN DSK2 OR SHE4 OR YMR276W OR YMR021.02.  
 OS *Saccharomyces cerevisiae* (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.  
 NCBI\_TaxID=4932;  
 RN [1]  
 RX STRAIN=S288c;  
 RC STRAIN=S288c;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX PubMed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;  
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome  
 RT XIII."  
 RL Nature 387:90-93 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX PubMed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;  
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome  
 RT XIII."  
 RL Nature 387:90-93 (1997).  
 CC -1- FUNCTION: INVOLVED, WITH RAD23 IN SPINDLE POLE BODY DUPLICATION.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: TO S.POMBE SPAC26A3.16.  
 CC -1- SIMILARITY: Contains 1 ubiquitin-like domain.  
 CC -1- SIMILARITY: Contains 1 UBA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L40587; AAB07267.1; --  
 DR PIR: S54583; S54583.  
 DR HSP: Q15843; INDD.  
 DR SGD: S0004889; DSK2.  
 DR InterPro: IPR006636; ST11.  
 DR InterPro: IPR000449; UBA domain.  
 DR InterPro: IPR006626; Ubiquitin.  
 DR Pfam: PF00627; UBA; 1.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS: PR00348; UBIQUITIN.  
 DR SMART: SM00727; ST11; 2.  
 DR SMART: SM00165; UBA; 1.  
 DR SMART: SM00213; UBO; 1.  
 DR PROSITE: PS50030; UBA; 1.  
 DR PROSITE: PS500299; UBIQUITIN\_1; 1.  
 DR PROSITE: PS50053; UBIQUITIN\_2; 1.  
 KW Nuclear protein.  
 FT DOMAIN 1 76 UBIQUITIN-LIKE.  
 FT DOMAIN 327 371 UBA.  
 FT CONFLICT 109 109 R -> A (IN REF. 2).  
 FT CONFLICT 296 296 R -> A (IN REF. 2).  
 SQ SEQUENCE 373 AA; 39516 MW; 25EDF82B9DB67DF6 CRC64;





RX MEDLINE=75156547; PubMed=124018;  
 RA Schlesinger D.H., Goldstein G.;  
 RT "Hybrid tropoin reconstituted from vertebrate and arthropod  
 subunits.";  
 RL Nature 255:423-424 (1975).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Mouse; STRAIN=C57BL/6;  
 RX MEDLINE=90245601; PubMed=2159627;  
 RA Finch J.S., Bonham K., Krieg P., Bowden G.T.;  
 RT "Murine polyubiquitin mRNA sequence.";  
 RL Nucleic Acids Res. 18:1907-1907 (1990).  
 RN [8]  
 RP SEQUENCE.  
 RC SPECIES=Rat; STRAIN=Wistar; TISSUE=Duoenum;  
 RX MEDLINE=94304928; PubMed=8031840;  
 RA Hubbard M.J., Carne A.;  
 RT "Differential feeding-related regulation of ubiquitin and  
 calbindin9kDa in rat duoenum.";  
 RL Biochim. Biophys. Acta 1200:191-196 (1994).  
 RN [9]  
 RP SEQUENCE FROM N.A. (UBA52).  
 RC SPECIES=Rat; STRAIN=Sprague-Dawley;  
 RX MEDLINE=96011832; PubMed=7488009;  
 RA Chan Y.-L., Suzuki K., Wool I.G.;  
 RT "The carboxyl extensions of two rat ubiquitin fusion proteins are  
 ribosomal proteins S27a and L40.";  
 RL Biochem. Biophys. Res. Commun. 215:682-690 (1995).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Bovine; TISSUE=Semenal vesicle;  
 RX MEDLINE=93176814; PubMed=8382528;  
 RA Wempe F., Scheit K.H.;  
 RT "Characterization of a full-length cDNA encoding a bovine four  
 tandem-repeat ubiquitin.";  
 RL Biochim. Biophys. Acta 1172:209-211 (1993).  
 RN [11]  
 RP SEQUENCE OF 1-74.  
 RC SPECIES=Bovine;  
 RX MEDLINE=75205496; PubMed=1170880;  
 RA Schlesinger D.H., Goldstein G., Niall H.D.;  
 RT "The complete amino acid sequence of ubiquitin, an adenylate cyclase  
 stimulating polypeptide probably universal in living cells.";  
 RL Biochemistry 14:2214-2218 (1975).  
 RN [12]  
 RP SEQUENCE OF 1-50.  
 RC SPECIES=Bovine;  
 RX MEDLINE=81062406; PubMed=6254502;  
 RA Hamilton J.W., Rouse J.B.;  
 RT "The biosynthesis of ubiquitin by parathyroid gland.";  
 RL Biochem. Biophys. Res. Commun. 96:114-120 (1980).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Pig;  
 RX MEDLINE=88054463; PubMed=2824145;  
 RA Einspanier R., Sharma H.S., Scheit K.H.;  
 RT "An mRNA encoding poly-ubiquitin in porcine corpus luteum:  
 identification by cDNA cloning and sequencing.";  
 RL DNA 6:395-400 (1987).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=C.porcillus; TISSUE=Spleen;  
 RA Tsukagoshi N.;  
 RT "Ascorbate-dependent expression of ubiquitin genes in guinea pig.";  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=C.griseus; STRAIN=V79; TISSUE=Lung fibroblast;  
 RX MEDLINE=92223105; PubMed=1314094;  
 RA Neno M., Mita K., Ichimura S.;  
 RT "Evolutionarily conserved structure of the 3' non-translated region  
 of a Chinese hamster polyubiquitin gene.";  
 RL Biochim. Biophys. Acta 1130:247-252 (1992).

RN [16]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=C.longicaudatus; STRAIN=V79; TISSUE=Lung fibroblast;  
 RX MEDLINE=89160248; PubMed=2537950;  
 RA Fornace A.J. Jr., Alamo I. Jr., Hollander M.C., Lamoreaux E.;  
 RT "Ubiquitin mRNA is a major stress-induced transcript in mammalian  
 cells.";  
 RL Nucleic Acids Res. 17:1215-1230 (1989).  
 RN [17]  
 RP SEQUENCE.  
 RC SPECIES=Rabbit; TISSUE=Brain;  
 RX MEDLINE=93149997; PubMed=1337207;  
 RA Wajih N., Siddiqi A.R., Kaiser R., Persson B., Zaidi Z.H.,  
 RA Joernvall H.;  
 RT "Structural characterization of rabbit brain ubiquitin.";  
 RL Protein Seq. Data Anal. 5:31-32 (1992).  
 RN [18]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Chicken;  
 RX MEDLINE=85213498; PubMed=2987683;  
 RA Bond U., Schlesinger M.J.;  
 RT "Ubiquitin is a heat shock protein in chicken embryo fibroblasts.";  
 RL Mol. Cell. Biol. 5:949-956 (1985).  
 RN [19]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Chicken;  
 RX MEDLINE=87089821; PubMed=3025663;  
 RA Bond U., Schlesinger M.J.;  
 RT "The chicken ubiquitin gene contains a heat shock promoter and  
 expresses an unstable mRNA in heat-shocked cells.";  
 RL Mol. Cell. Biol. 6:4602-4610 (1986).  
 RN [20]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Chicken; TISSUE=Spermatid;  
 RX MEDLINE=89098351; PubMed=2850543;  
 RA Mezquita J., Pau M., Mezquita C.;  
 RT "cDNA encoding a chicken ubiquitin-fusion protein identical to the  
 corresponding human protein.";  
 RL Nucleic Acids Res. 16:11838-11838 (1988).  
 RN [21]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=X.laevis;  
 RX MEDLINE=85048942; PubMed=6209017;  
 RA Dworkin-Rastl E., Shrutkowski A., Dworkin M.B.;  
 RT "Multiple ubiquitin mRNAs during Xenopus laevis development contain  
 tandem repeats of the 76 amino acid coding sequence.";  
 RL Cell 39:321-325 (1984).  
 RN [22]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RC SPECIES=Human; PubMed=3041007;  
 RX MEDLINE=87311725; PubMed=3041007;  
 RA Vijay-Kumar S., Bugg C.E., Cook W.J.;  
 RT "Structure of ubiquitin refined at 1.8-A resolution.";  
 RL J. Mol. Biol. 194:531-544 (1987).  
 RN [23]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RC SPECIES=Human; PubMed=8166633;  
 RX MEDLINE=94220022; PubMed=8166633;  
 RA Ramage R., Green J., Muir T.W., Ogunjobi O.M., Love S., Shaw K.;  
 RT "Synthetic, structural and biological studies of the ubiquitin system:

Query Match 6.4%; Score 122.5; DB 1; Length 76;  
 Best Local Similarity 40.5%; Pred. No. 0.085;  
 Matches 30; Conservative 14; Mismatches 27; Indels 3; Gaps 1;  
 Qy 1 MKLTVKTLKGTHTFEIRYQPNNDTMAVKNTIEIQKDSYPWGQQLIFNGKVLKDESTLE 60  
 Db 1 MQIFVKLTGTITILEVEPSDTIENVK---AKIQKEGIPDQQLIFAGKLEDCRTLS 57  
 Qy 61 ENKYNEDGFLVVML 74  
 Db 58 DYNIQKESTLHLVL 71

```

RESULT 11
RFX1 HUMAN STANDARD; PRT; 979 AA.
ID AC P22670;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE MHC class II regulatory factor RFX1 (RFX) (Enhancer factor C) (EF-C).
GN RFX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RP MEDLINE=91071581; PubMed=2253877;
RA Reith W., Sanchez-Herrero C., Kober M., Silacci P., Berte C.,
RA Barras E., Mach B.;
RT "MHC class II regulatory factor RFX has a novel DNA-binding domain
RT and a functionally independent dimerization domain.";
RL Genes Dev. 4:1528-1540(1990).
[2]
RP IDENTITY BETWEEN RFX1 AND EF-C.
RP MEDLINE=94019311; PubMed=8413236;
RA Siegrist C.A., Durand B., Emery P., David E., Hearing P., Mach B.,
RA Reith W.;
RT "RFX1 is identical to enhancer factor C and functions as a
RT transactivator of the hepatitis B virus enhancer.";
RL Mol. Cell. Biol. 13:6375-6384(1993).
[3]
RP BINDING TO RPL30 PROMOTER.
RP MEDLINE=94040774; PubMed=8224874;
RA Saitany G., Perry R.P.;
RT "Transcription factor RFX1 helps control the promoter of the mouse
RT ribosomal protein-encoding gene rpl30 by binding to its alpha
RT element.";
RL Gene 132:279-283(1993).
[4]
RP SHOWS THAT BLS II IS NOT DUE TO RFX1.
RP MEDLINE=92375076; PubMed=1508204;
RA Sanchez-Herrero C., Reith W., Silacci P., Mach B.;
RT "The DNA-binding defect observed in major histocompatibility complex
RT class II regulatory mutants concerns only one member of a family of
RT complexes binding to the X boxes of class II promoters.";
RL Mol. Cell. Biol. 12:4076-4083(1992).
CC -1- FUNCTION: REGULATORY FACTOR ESSENTIAL FOR MHC CLASS II GENE
CC EXPRESSION. BINDS TO THE X BOXES OF MHC CLASS II GENES. ALSO BINDS
CC TO AN INVERTED REPEAT (ENH1) REQUIRED FOR HEPATITIS B VIRUS GENES
CC EXPRESSION AND TO THE MOST UPSTREAM ELEMENT (ALPHA) OF THE RPL30
CC PROMOTER.
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE RFX FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announcement/)
or send an email to license@isb-sib.ch).
-----
ENBL; X58964; CAA01730.1; --
ENBL; A20498; CAA01506.1; --
PIR; A35913; A35913.
PDB; 1DP7; 06-MAR-00.
DR TRANSFAC; T00909; --
DR TRANSFAC; T01673; --
DR Genew; HGNC:9982; RFX1.
DR MIM; 600006; --
DR GO; GO:0003705; P:RNA polymerase II transcription factor acti. .; TAS.
DR GO; GO:0006955; P:immune response; TAS.

```



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D23662; BAA04889.1; -;  
PDB; 1NDJ; 23-FEB-99.  
Genew; HGNC:7732; NEDD8.  
MIM; 603171; -;  
GO; GO:0005634; C:nucleus; TAS.  
GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.  
GO; GO:0006464; P:protein modification; TAS.  
GO; GO:0006508; P:proteolysis and peptidolysis; TAS.  
GO; GO:0006511; P:ubiquitin-dependent protein catabolism; TAS.  
InterPro; IPR000626; Ubiquitin.  
Pfam; PF00240; ubiquitin; 1.  
PRINTS; PR00348; UBIQUITIN.  
SMART; SM00213; UBQ; 1.  
PROSITE; PS00299; UBIQUITIN\_1; 1.  
PROSITE; PS50053; UBIQUITIN\_2; 1.  
Ubl conjugation pathway; 3D-structure.  
FT STRAND 2 7  
TURN 8 9  
FT STRAND 12 16  
TURN 19 20  
FT STRAND 22 22  
TURN 23 34  
FT HELIX 38 40  
TURN 41 45  
FT STRAND 46 47  
TURN 48 49  
FT STRAND 52 53  
TURN 55 55  
FT STRAND 56 60  
TURN 63 64  
FT STRAND 66 71  
TURN 74 75  
SQ SEQUENCE 81 AA; 9072 MW; DC2FE102BE4725D2 CRC64;

Query Match 6.4%; Score 121.5; DB 1; Length 81;  
Best Local Similarity 36.0%; Pred. No. 0.11;  
Matches 27; Conservative 16; Mismatches 29; Indels 3; Gaps 1;  
QY 1 MKLTVKLTGKTHFEIRVQNDTIMAVKKNIEIQGKDSYPWGQQLLIFNGKVLKDESTLE 60  
DB 1 MLIKVKLTGKEIDIEPTDKVERKERVEE---KEGIPQQRLLYSGKQWMDKATAA 57  
QY 61 ENKVNEDGFLVWMLS 75  
DB 58 DYKILGGSVLHLVLA 72

RESULT 13  
NED8\_MOUSE  
ID NED8\_MOUSE STANDARD; PRT; 81 AA.  
AC P29595;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Ubiquitin-like protein NEDD8.  
GN NEDD8 OR NEDD-8.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92328780; PubMed=1378265;  
RA Kumar S., Tomooka Y., Noda M.;  
RT "Identification of a set of genes with developmentally down-regulated expression in the mouse brain."  
RL Biochem. Biophys. Res. Commun. 185:1155-1161(1992).  
RN [2]

RP REVISIONS, SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=93371434; PubMed=8395831;  
RT Kumar S., Yoshida Y., Noda M.;  
RL "Cloning of a cDNA which encodes a novel ubiquitin-like protein.";  
CC Biochem. Biophys. Res. Commun. 195:393-399(1993).  
CC -!- FUNCTION: ACTIVATED BY AN E1-LIKE COMPLEX, CONSISTING OF APP-B1  
CC AND UBA3 AND THEN LINKED TO THE E2-LIKE ENZYME, UBC12. THE MAJOR  
CC TARGET PROTEIN MODIFIED BY NEDD8 IS CULLIN-4A.  
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE DURING THE EMBRYONIC  
CC DEVELOPMENT AND DIFFERENTIATION OF THE CENTRAL NERVOUS SYSTEM. MAY  
CC PLAY AN ESSENTIAL ROLE IN EUKARYOTIC CELLULAR METABOLISM.  
CC -!- TISSUE SPECIFICITY: Ubiquitous.  
CC -!- DEVELOPMENTAL STAGE: DOWN-REGULATED DURING THE DEVELOPMENT OF  
CC BRAIN.  
CC -!- SIMILARITY: BELONGS TO THE UBIQUITIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D10918; BAA01719.1; -;  
FIR; JN0710; JN0710.  
HSSP; Q15843; 1NDD.  
MGD; MGI:97301; Nedd8.  
InterPro; IPR000626; Ubiquitin.  
Pfam; PF00240; ubiquitin; 1.  
PRINTS; PR00348; UBIQUITIN.  
SMART; SM00213; UBQ; 1.  
PROSITE; PS00299; UBIQUITIN\_1; 1.  
PROSITE; PS50053; UBIQUITIN\_2; 1.  
Ubl conjugation pathway.  
KW Ubl conjugation pathway.  
SQ SEQUENCE 81 AA; 8972 MW; DC339102BE4725D2 CRC64;

Query Match 6.4%; Score 121.5; DB 1; Length 81;  
Best Local Similarity 36.0%; Pred. No. 0.11;  
Matches 27; Conservative 16; Mismatches 29; Indels 3; Gaps 1;  
QY 1 MKLTVKLTGKTHFEIRVQNDTIMAVKKNIEIQGKDSYPWGQQLLIFNGKVLKDESTLE 60  
DB 1 MLIKVKLTGKEIDIEPTDKVERKERVEE---KEGIPQQRLLYSGKQWMDKATAA 57  
QY 61 ENKVNEDGFLVWMLS 75  
DB 58 DYKILGGSVLHLVLA 72

RESULT 14  
UBIQ\_EIMBO  
ID UBIQ\_EIMBO STANDARD; PRT; 76 AA.  
AC P46574;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Ubiquitin.  
OS Eimeria bovis.  
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
OC Eimeria.  
OX NCBI\_TaxID=5803;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93149194; PubMed=8426605;  
RA Abrahamson M.S., Clark T.G., Mascolo P., Speer C.A., White M.W.;  
RT "Developmental gene expression in Eimeria bovis."  
RL Mol. Biochem. Parasitol. 57:1-14(1993).  
CC -!- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
CC BIOGENESIS.



```

KW Nuclear protein; Polyprotein.
FT SITE 48
FT BINDING 76 76
FT SEQUENCE 76 AA; 8549 MW; 52648497FPD8E47 CRC64;
SQ

Query Match 6.2%; Score 118.5; DB 1; Length 76;
Best Local Similarity 39.2%; Pred. No. 0.15;
Matches 29; Conservative 15; Mismatches 27; Indels 3; Gaps 1;

QY 1 MKLTVTKLGTGTHFEIRVQPNQDTIMAVKNTIEETQCKDSYPAGQQLLIFNGKVLKDESTLE 60
Db 1 MQIFVKTLTKKTIALEVEPSDTIENVK---AKIQKKEGIPPDQORLIFAGKQLEGRRLS 57
QY 61 ENKVNEDGFLVVMVL 74
Db 58 DYNIQKESTLHLVL 71

Search completed: December 17, 2003, 06:17:19
Job time : 11.4735 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:16:40 ; Search time 37.6093 Seconds  
(without alignments)  
2524.996 Million cell updates/sec

Title: US-09-805-550-4  
Perfect score: 1910  
Sequence: 1 MKLTVTLKGTHTPEIRVQPN.....CDRNEBLAANYLLEHAGEED 368

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTRMBL 23.1

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_podent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1169.5	61.2	365	10 Q94CE9	Q94ce9 arabidopsis
2	1169.5	61.2	371	10 Q8LA46	Q8la46 arabidopsis
3	1153.5	60.4	379	10 O03991	O03991 daucus caro
4	1043.5	54.6	367	10 Q9MA10	Q9ma10 arabidopsis
5	1019	53.4	389	10 Q9STA6	Q9sta6 lycopersico
6	978	51.2	419	10 Q9W887	Q9w887 arabidopsis
7	954	49.9	382	10 O03990	O03990 daucus caro
8	931.5	48.8	378	10 Q9FF16	Q9ff16 arabidopsis
9	930.5	48.7	378	10 Q9AC35	Q9ac35 arabidopsis
10	887.5	46.5	392	10 Q40742	Q40742 oryza sativ
11	661.5	34.6	246	10 Q9S9L8	Q9s9l8 arabidopsis
12	621.5	32.5	409	4 Q8WUB0	Q8wub0 homo sapien
13	599	31.4	362	11 Q8CAP3	Q8cap3 mus musculu
14	490.5	25.7	414	5 Q9XZE0	Q9xze0 drosophila
15	487.5	25.5	414	5 Q9V3W9	Q9v3w9 drosophila
16	484	25.3	341	5 Q97135	Q97135 dictyosteli

17	431	22.6	296	11 Q8BRA9	Q8bra9 mus musculu
18	424	22.2	113	10 Q9SA20	Q9sa20 arabidopsis
19	410.5	21.5	389	5 Q81JS8	Q81js8 plasmodium
20	405	21.2	372	5 Q23451	Q23451 caenorhabdi
21	405	21.2	748	6 P79370	P79370 oryctolagus
22	382	20.0	343	5 Q81MB7	Q81mb7 drosophila
23	375	19.6	290	5 Q9VCD5	Q9vcd5 drosophila
24	215	11.3	523	5 Q9NIF3	Q9nif3 dictyosteli
25	214	11.2	575	10 Q9FWF5	Q9fwf5 oryza sativ
26	204.5	10.7	142	10 Q9LF19	Q9lf19 arabidopsis
27	201	10.5	551	10 Q94C51	Q94c51 arabidopsis
28	201	10.5	551	10 Q9S1I8	Q9s1i8 arabidopsis
29	194.5	10.2	538	10 Q9S1I9	Q9s1i9 arabidopsis
30	193	10.1	65	10 Q9SCA8	Q9sc8 lycopersico
31	180.5	9.5	582	11 Q9QZM1	Q9qzm1 mus musculu
32	176	9.2	589	4 Q9H0T8	Q9h0t8 homo sapien
33	176	9.2	589	4 Q9UMX0	Q9umx0 homo sapien
34	176	9.2	589	4 Q81XS9	Q81xs9 homo sapien
35	174	9.1	595	6 Q95M59	Q95m59 bos taurus
36	171	9.0	601	4 Q9NRR5	Q9nrr5 homo sapien
37	170.5	8.9	582	11 Q9JJP9	Q9jjp9 rattus norv
38	170.5	8.9	596	11 Q9N8B8	Q9n8b8 mus musculu
39	169	8.8	314	4 Q9H8R7	Q9h8r7 homo sapien
40	167.5	8.8	555	11 Q8C7T4	Q8c7t4 mus musculu
41	167.5	8.8	582	11 Q8R317	Q8r317 mus musculu
42	167	8.7	638	11 Q9QZM0	Q9qzm0 mus musculu
43	165	8.6	589	4 Q9H3R4	Q9h3r4 homo sapien
44	163.5	8.6	502	5 Q18672	Q18672 caenorhabdi
45	160	8.4	548	10 Q8LSP9	Q8lsp9 oryza sativ

## ALIGNMENTS

## RESULT 1

Q94CE9 PRELIMINARY; PRT; 365 AA.  
ID AC Q94CE9;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative RAD23 protein.  
GN F20B17.8.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsie.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,  
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,  
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,  
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,  
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,  
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,  
RT "Full length cDNA of gene F20B17.8 (GI:7715605).";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
RP [2]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,  
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,  
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,  
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,  
RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Full length cDNA of gene F20B17.8 (GI:7715605).";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY034912; AAK59419.1; -;  
DR EMBL; AY063103; AAL34277.1; -;

```

DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STI1.
DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR SMART; SM00727; STI1; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS00053; UBIQUITIN_2; 1.
SQ SEQUENCE 365 AA; 39157 MW; 4FED9EC59B467745 CRC64;

Query Match 61.2%; Score 1169.5; DB 10; Length 365;
Best Local Similarity 65.1%; Pred. No. 4.6e-75;
Matches 246; Conservative 43; Mismatches 66; Indels 23; Gaps 9;

Qy 1 MKLTIVTKLKGTHFEIRVOPNDTIMAVKKNIEIQKDSYPMGQQLLIHNGKVLKDESTLE 60
Db 1 MKLTIVTKLKGSHFEIRVLPSTTIMAVKKNIEDSQKDNYPGQQLLIHNGKVLKDESTLV 60

Qy 61 ENKVNEDGLVVLMSKGTSGTSSQ--HSNTPATROAPPLEAQPOAPPPVAPITT 118
Db 61 ENKVTGEGFLVVLMSKSGSAGASQVPSATTSSTKPAAP--STTQSSPVPASPIPA 118

Qy 119 SOPEGLPAQAPNTHNAASLLSGRNVDTIINQLMEMGGGWDKDKVORALRAA 178
Db 119 QEQ---PAAQTDITYGQAASLTLSGSSLEQVMQVQIIMEMGGGWDKKTETVTRALRAA 175

Qy 179 AVEYLISGIPVTAETAVP-----IGGQGANTTDRAPTGEAGLSGIPNTAPLDLPQAGSN 233
Db 176 AVDYLYSGIPQTAETAVPPEAQIAGSGA-----APVAPA--SGGPNSSPLDLFPQETVA 228

Qy 234 AGGAGGGGLDFLRNPNQAVREMVHTNPQILQPMVLVELSKNPQILRLIENHDEFLQ 293
Db 229 AAGSGDLGLTFLRNNDQQLRTWVHSNPNQILQPMVLQELGKQNPQLRLIENHDEFLQ 288

Qy 294 LINEPFEQ--GGGDFLDQPEDEMPHAI SVTPEEQEATGRLESNGFDRARVIEAFLACDR 351
Db 289 LVNERPYEGSDGDMFDQPEQ--EMPHAINVTPEAQEAIQRLGAMGFDRALVIEAFLACDR 347

Qy 352 NEELAANYLLEHAGE-ED 368
Db 348 NEELAANYLLENSGDFED 365

RESULT 2
Q8LA46 PRELIMINARY; PRT; 371 AA.
AC Q8LA46;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE DNA repair protein RAD23, putative.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
annotation."
RL Genome Biol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY088037; AAM65583.1; -.

```

```

DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STI1.
DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR SMART; SM00727; STI1; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS00053; UBIQUITIN_2; 1.
SQ SEQUENCE 371 AA; 39747 MW; 081493086EA97657 CRC64;

Query Match 61.2%; Score 1169.5; DB 10; Length 371;
Best Local Similarity 64.3%; Pred. No. 4.7e-75;
Matches 247; Conservative 43; Mismatches 65; Indels 29; Gaps 9;

Qy 1 MKLTIVTKLKGTHFEIRVOPNDTIMAVKKNIEIQKDSYPMGQQLLIHNGKVLKDESTLE 60
Db 1 MKLTIVTKLKGSHFEIRVLPSTTIMAVKKNIEDSQKDNYPGQQLLIHNGKVLKDESTLV 60

Qy 61 ENKVNEDGLVVLMSKGTSGTSSQHSN-----TPATROAPPLEAQPOAPPP 112
Db 61 ENKVTGEGFLVVLMSKSGSAGASQVPSATTSSTKPAAP--STTQSSPVP 118

Qy 113 VAPITTSQPEGLPAQAPNTHNAASLLSGRNVDTIINQLMEMGGGWDKDKVORALRAA 172
Db 119 ASPIPAEQ---PAAQTDITYGQAASLTLSGSSLEQVMQVQIIMEMGGGWDKKTETVTRALRAA 175

Qy 173 YNNPERAVELYSGIPVTAETAVP-----IGGQGANTTDRAPTGEAGLSGIPNTAPLDLF 227
Db 176 YNNPERAVDYLYSGIPQTAETAVPPEAQIAGSGA-----APVAPA--SGGPNSSPLDLF 228

Qy 228 PQGASNAGGAGGGGLDFLRNPNQAVREMVHTNPQILQPMVLVELSKNPQILRLIEN 287
Db 229 PQETVAAGSGDLGLTFLRNNDQQLRTWVHSNPNQILQPMVLQELGKQNPQLRLIEN 288

Qy 288 HDEFQLLINEPFEQ--GGGDFLDQPEDEMPHAI SVTPEEQEATGRLESNGFDRARVIEA 345
Db 289 QAEFLQLVNERPYEGSDGDMFDQPEQ--EMPHAINVTPEAQEAIQRLGAMGFDRALVIEA 347

Qy 346 FLACDRNEELAANYLLEHAGE-ED 368
Db 348 FLACDRNEELAANYLLENSGDFED 371

RESULT 3
O03991 PRELIMINARY; PRT; 379 AA.
AC O03991;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RAD23 protein, isoform II.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=W001C;
RX MEDLINE=98345997; PubMed=9681019;
RA Sturm A., Leinhard S.;
RT "Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in
Yeast."
RL Plant J. 13:815-821(1998).
DR EMBL; Y12014; CAA72742.1; -.
DR HSP; P54725; LDV0.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STI1.
DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.

```

```

DR Pfam; PF00240; ubiquitous; 1.
DR SMART; SM00727; still; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBA; 1.
DR TIGRfam; TIGR00601; rad23; 1.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
SQ SEQUENCE 379 AA; 40530 MW; B26697B439CC5929 CRC64;

Query Match
Best Local Similarity 60.4%; Score 1153.5; DB 10; Length 379;
Matches 240; Conservative 45; Mismatches 81; Indels 15; Gaps 9;

QY 1 MKLTVTLKGTHTFIRVQPNDDTMAVKNIIEIQKDSYPMGQQLIFNGKVLKDESTLE 60
DB 1 MKLTVTLKGTSHFEIRVLPNDTMAVKNIIEIQKDSYPMGQQLIFNGKVLKDESTLE 60

QY 61 ENKVNEDGFLVVLMSKGTSGTSSQHSNTPTATPQAP---PLEAPQAPQPPVAPIT 117
DB 61 ESKISDGFLLVVLMSKGTSGTSSQHSNTPTATPQAP---PLEAPQAPQPPVAPIT 120

QY 118 TSQEG--LPAQAP-NTHDAAASNLGRNVDTIINQLMEMGGSGDKVKORALRAAYN 174
DB 121 T-VPEAPLSPAPAPSDTYGEAASNVVAGSNLEQTIQHIMDMGGMDTNVSEALRAAYN 179

QY 175 NPERAVYLYSGIPVTAEIAPVIG---GQANTTDRA--PTGBAGLS-GIPNTPADLPLP 228
DB 180 NPERAVDLYSGIPVMAEAAVPSVHFQDQINAGNNAISDNGVAGAAPGAPNSLPLNMFF 239

QY 229 QGASNAGGAGGGLPDLRNNPOFQAVREWHNTNPQILOPMLVELSKONPQILRIENH 288
DB 240 QETLSGVTGAGLSLEFLRNNPOFQILRSVQRNPOILOPMLLELGKONPQILRIQIEH 299

QY 289 DEFLQLNPFEGEGEDFLDQPEDEMPHAISVTPPEQBAIGRLSMGDFRVIARFLA 348
DB 300 EEFQLNPFVEASEGDMFDQPEQD-VPEITVTADQAEIRLEANGDFRGLVIEARFLA 358

QY 349 CORNEELAVNYLLEHAGE-ED 368
DB 359 CORNEELAVNYLLEHAGE-ED 379

RESULT 4
Q9MA10 PRELIMINARY; PRT; 367 AA.
AC Q9MA10;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE F20B17.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B.,
RA Lenz C., Li J., Liu A., Liu J., Liu J., Liu J., Liu J., Liu J.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F20B17 from chromosome
RT I.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;

```

```

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu J., Liu J., Liu J.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010793; AAF68123.1; -
DR HSSP; P54725; IDV0.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBA; 1.
DR TIGRfam; TIGR00601; rad23; 1.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
SQ SEQUENCE 367 AA; 39790 MW; C6D01134B5C69CE6 CRC64;

Query Match
Best Local Similarity 54.6%; Score 1043.5; DB 10; Length 367;
Matches 232; Conservative 39; Mismatches 73; Indels 47; Gaps 11;

QY 1 MKLTVTLKGTHTFIRVQPNDDTMAVKNIIEIQKDSYPMGQQLIFNGKVLKDESTLE 60
DB 1 MKLTVTLKGTSHFEIRVLPNDTMAVKNIIEIQKDSYPMGQQLIFNGKVLKDESTLE 60

QY 61 ENKVNEDGFLVVLMSKGTSGTSSQHSNTPTATPQAP---PLEAPQAPQPPVAPIT 113
DB 61 ENKVTBEGFLVVLMSKSGSAGASVQCVRLLLFHSLFP-----LPHRLSLSTY 111

QY 114 APITT-----SQEGLPAQAPNTHDAAASNLGRNVDTIINQLMEMGGSGDKVK 165
DB 112 NPVFTCSCTYSCTTGTG----TDTYGMAASTLVSGSSLEQMVQIIMEMGGSGDKV 166

QY 166 QALRAAYNNPERAVYLYSGIPVTAEIAPV-----IGQGANTTDRAPTGBAGLSGIPN 220
DB 167 TRALRAAYNNPERAVDLYSGIPVTAEIAPVPEAQIAGSGA-----APVAPA--SGGN 219

QY 221 TAPDLFPQASNAGGAGGGLPDLRNNPOFQAVREWHNTNPQILOPMLVELSKONPQI 280
DB 220 SPDLFPQETVAAAGSGDLGTLPLRNNQVAIL--TISAFSLNCEPMLQSLGKONPOL 277

QY 281 LRLIENHDEFLQLNPFEG--GEGDFLDQPEDEMPHAISVTPPEQBAIGRLSMGDF 338
DB 278 LRLIQENQAEFLQVNEPEYSGDGEDMFDQPEQ-EMPHAINVTAEQAEIQRLEANGFD 336

QY 339 RARVTEAFIACDRNEELAVNYLLEHAGE-ED 368
DB 337 RALVTEAFIACDRNEELAVNYLLEHAGE-ED 367

RESULT 5
Q9STA6 PRELIMINARY; PRT; 389 AA.
ID Q9STA6;
AC Q9STA6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RAD23 protein.
GN RAD23.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=cv. West virginia 106; TISSUE=Fruit;  
 RA Lemaire-Chamley M., Petit J., Raymond P., Chevallier C.;  
 RT "Analysis of gene expression during early tomato fruit development by  
 mRNA differential display.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR HMBL; AJ243875; CAB51544.1; -.  
 DR HSP; P54725; IDVO.  
 DR InterPro: IPR004806; Rad23.  
 DR InterPro: IPR006636; SRI1.  
 DR InterPro: IPR000449; UBA\_domain.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00627; UBA; 2.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR SMART; SM00727; STI1; 1.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS50053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 389 AA; 41508 MW; D2BAEDF0FE70778A CRC64;  
 Query Match 53.4%; Score 1019; DB 10; Length 389;  
 Best Local Similarity 53.8%; Pred. No. 2.5e-64;  
 Matches 214; Conservative 47; Mismatches 97; Indels 40; Gaps 6;  
 QY 1 MKLTWKTKGTHFIRVOPNDTIMAVKQIEEIQKDSYPMGQQLLIFNGKVLKDESTLE 60  
 DB 1 MKIFVKTKGTHFIEVKPEDSVADVKNIESVQGVYPAQQLIHQGVKLDKTTTLE 60  
 QY 61 ENKYNEDGFLVVMLSKGTSGTSS-----SQHSNTPATRQAPPL 102  
 DB 61 ENKVAENSFVIMLSKQKVS-STGTSSISAALNTAQDGTGTDQARQITTFQATAALPQ 119  
 QY 103 EAPQAPQPPVAPITTSQPEGLPAQPNTHNAAISLLSGHNVDTIINOLMEMGGGWDK 162  
 DB 120 SASESAPTPAPVPAAS-----SVTDVYDQASNLVAGSNLETTVQQLDMGGGWDK 172  
 QY 163 DKVORALRAANNPERAVEYLYSGIPVTAET-----AVPIGGQGANTTDRAPT 210  
 DB 173 DTVRALRAANNPERAVDYLYSGIPEQTEIPPARAPAAVATPASAQAINPAQDAS 232  
 QY 211 GEAGLSGIPNTAPLDLPQGNAGGGAGGGGLFLRNPNQFQAVREMVHTNPQILQPLM 270  
 DB 233 QLAVPSSGPNANPLDLPQGLTNAGSNAGAGNLDPLRNSPQFQALRAMVQANPQILQPLM 292  
 QY 271 VELSKQNPQILRLIENHDEFLLQNEPPEGGDFLDQPEDEMPHAISVTPEQEAIG 330  
 DB 293 QELGKQNPMLRLIQEHQDFLRLINEPVE-GEQNVLGQ-TAGATPQAVTVTPEREAE 350  
 QY 331 RLEMGFDRARVIEAFLACDRNEELAANYLLEHAGEED 368  
 DB 351 RLEMGFDRALVLEVYFACNKEELAANYLLDHLHEFD 388  
 RESULT 6  
 QSM897 PRELIMINARY; PRT; 419 AA.  
 ID Q9M887  
 AC Q9M887;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Putative RAD23 (AT3G02540/F16B3\_17).  
 GN F16B3.17.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsia.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,  
 RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;

RT "Arabidopsis thaliana chromosome III BAC F16B3 genomic sequence.";  
 Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Gibson H.A., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,  
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,  
 RA Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis cDNA clones."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Shinn P., Banh J., Bowser L.,  
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis ORF clones."  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC021640; AAF32461.1; -.  
 DR EMBL; AV039562; AAK62617.1; -.  
 DR EMBL; AY113034; AAM47342.1; -.  
 DR HSP; P54725; IDVO.  
 DR InterPro: IPR002965; P-rich\_extensn.  
 DR InterPro: IPR004806; Rad23.  
 DR InterPro: IPR000449; UBA\_domain.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00627; UBA; 2.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS50053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 419 AA; 44247 MW; CAAL3BC4FFEB1E25 CRC64;  
 Query Match 51.2%; Score 978; DB 10; Length 419;  
 Best Local Similarity 50.6%; Pred. No. 2.2e-61;  
 Matches 213; Conservative 58; Mismatches 90; Indels 60; Gaps 12;  
 QY 1 MKLTWKTKGTHFIRVOPNDTIMAVKQIEEIQKDSYPMGQQLLIFNGKVLKDESTLE 60  
 DB 1 MKIFVKTKGTHFIEVKPEDSVADVKNIESVQGVYPAQQLIHQGVKLDKTTTLE 60  
 QY 61 ENKYNEDGFLVVMLSKGT-----TSGSTGT-----SSQHS-----NTPATRQAP--- 100  
 DB 61 ENKVAENSFVIMMNSKPASAAASASACTSQAKSTPPSTQSPISQTPASVSAPVAP 120  
 QY 101 -PLEAPQAPQPPVAP-----ITTSQPEGLPA-----QAP-----NTHDNAASN 138  
 DB 121 APTRPPTPTPTPAPVAATETVTTPPEVPATISSSTPAPDSAPVGSQGVYQAASN 180  
 QY 139 LLSGNVDTIINOLMEMGGGWDKQVORALRAANNPERAVEYLYSGIPVTAIEVPI- 197  
 DB 181 LAAGSNLESTIQQLDMGGGTWDRETVYALRAAFNPERAVEYLYTGTIEQAEVP-FVA 239  
 QY 198 -----GGQANTTDRAPTGEAGLSGIPNTAPLDLPQGNAGGGAGGGGLFLRNPNQF 252  
 DB 240 RPPASACQPNPAPQATQQAAPASGPNANPLDLPQGLPNVGNPAGTLDLFRNSQOF 299  
 QY 253 QAVREMVHTNPQILQPLMVELSKQNPQILRLIENHDEFLLQNEPPEGG--EGDFLDQ- 309  
 DB 300 QALRAMVQANPQILQPLMVELSKQNPMLRLIQHQAQDFLRLINEPPEVGGGSGNLLQCM 359  
 QY 310 -----PEDEMPHAISVTPEQEAIGRLSEMGFDRARVIEAFLACDRNEELAANYLLEHAG 365  
 DB 360 AAGMPQ-----PQATQVTHEEREALERLEANGFERALVLEVFACNKEELAANYLLDHMH 415  
 QY 366 E 366

Db 416 E 416

```

RESULT 7
O03990
ID O03990 PRELIMINARY; PRT; 382 AA.
AC O03990;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE RAD23, isoform 1.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=W001C;
RX MEDLINE=98345997; PubMed=9681019;
RA Sturm A., Leinhardt S.;
RT "Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in
RT yeast.";
RL Plant J. 13:815-821(1998).
DR EMBL; Y12013; CAA72741.1; -.
DR HSSP; P54725; 1DVO.
DR InterPro; IPR002965; P-rich_extensn.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STL.
DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR PRINTS; PR01217; PRICHTEXTENS.
DR SMART; SM00727; STL1; 1.
DR SMART; SM00165; UBA; 2.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS50053; UBQUITIN 2; 1.
SQ SEQUENCE 382 AA; 40344 MW; 27F02A2402F3DC2 CRC64;

Query Match 49.9%; Score 954; DB 10; Length 382;
Best Local Similarity 52.4%; Pred. No. 9.9e-60;
Matches 205; Conservative 52; Mismatches 98; Indels 36; Gaps 9;

QY 1 MKLTVTLLKGTHTFTRVQPNPTIMAVKNIEIQKDSYPRGQQLLIINGKVLKDESTLE 60
DB 1 MKIYVTKLKGQFEIQVNPDDSDVADKRSIETAQGAAYPAQQMLIYQGKVLKDGTTLL 60
QY 61 ENKVNEDGFLVVLMSKGTSGTGTSSQHSNTPATROAPPLAEPQAP- - - - -OPVVA- 114
DB 61 ENNVAENSVIIVLMSKSPSGEGSTTS- - - - -TAAAPKAPQTSAPSPVPAVQPPAST 116
QY 115 -PITTSQEGGLPAQAP- - - - -NTHNAASNLSSGRVDTTIINQLMEMGGSGWDK 163
DB 117 LPVPAPSPAPAPATAPIPSAAVGSSEANVYDASALLVAGSLLEGAIQIILDMGGGTWDR 176
QY 164 KQVRLRAAYNNPERAVELYSGIPVTAETIAVPI- - - - -GGGANTTDRAPTG- - -EAGLS 216
DB 177 TVIRIVRAAFNNPERAVELYSGIPEQAE-APPVAPSPSPGQAANPLDQPPAAQAPAPAS 235
QY 217 GIPNTAPLDLPQGNASAG- - - - -CGAGGGPLDLRNNPFOCAVREWHVTNPQIILQMLVELSK 275
DB 236 AGPNANPLDLFPQGLPDGMSNAGNLDLFTNQFQALRAWVQSNPQIILQMLQELGK 295
QY 276 QNPQILRIEENHIEFLQLLNEPEFGEGGDFDQPEDEMPHAI SVTPEEQEAIGRLESM 335
DB 296 QNPHLMRLIQHQADFLQLINEPNEGGENLJ- - - - -CHGQQAISVTPEDDAERLEAM 349
QY 336 GFDRARVTEAFACDRNEELAAANYLLEHAGE 366
DB 350 GFDRFLVLEVFVACNKEELAANYLLDHME 380

```

```

RESULT 8
O9FF16
ID O9FF16 PRELIMINARY; PRT; 378 AA.
AC O9FF16;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE DNA repair protein RAD23 homolog (Hypothetical protein)
DE (A5938470/At5g38470)
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyaajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT P1 clones.";
RL DNA Res. 4:215-230(1997).
RN [2]_
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayaashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cdna clones.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]_
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayaashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]_
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0(2002).
RN [5]_
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB005248; BAB09359.1; -.
DR EMBL; AY058196; AAL25609.1; -.
DR EMBL; AY081835; AAL87405.1; -.
DR EMBL; AY087564; AAM65106.1; -.
DR HSSP; P54725; 1DVO.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.

```

Wed Dec 17 06:49:44 2003

DR TIGRFAMs; TIGR00601; rad23; 1.  
DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 378 AA; 40066 MW; DD9B59153EA3B6A9 CRC64;

Query Match 48.8%; Score 931.5; DB 10; Length 378;  
Best Local Similarity 52.6%; Pred. No. 3.9e-58;  
Matches 200; Conservative 56; Mismatches 109; Indels 15; Gaps 8;

QY 1 MKLTVKTLKGTGHEIRVQPNDDTMAVKKNIEEIOGKDSYPMGQQLIFNGKVLKDESTLE 60  
DB 1 MKLIFVKTLSSGNEFEIEVKPADKVDKTAETVKGAE-YPAAKQMLIHQGVKLDSTLE 59  
QY 61 ENKVNEDGFLVVMLSKGKTS---GSTGTSSSHSNTPATRQAPPLEAP-QQAPQPPVAPI 116  
DB 60 ENNVVENSFVIMLSKTKASPSGASTASAPASATQPTVATPQVSAPTASVPVTSGTA 119  
QY 117 TTSQPEGLPAQAPNTHDAAASNLSSGRNVDTIINQLMGGGSGWDKQVORALRAAYNNP 176  
DB 120 TAAAPATAASVQTDVYQGAASNLVAGTTLESTVQQLIDMGSGSWDRDVTVVRALRAAFNNP 179  
QY 177 ERAVEYLISGIPVTAETI-----AVPIGGQGAN--TTDRAPTGEAGLSGIPNTAPLDLFPQ 229  
DB 180 ERAVEYLISGIPVTAETI-----AVPIGGQGAN--TTDRAPTGEAGLSGIPNTAPLDLFPQ 239  
QY 230 GASNAGGAGGGLDPLFRNNPQFOAVREVMVHTNPQILQPMVLVELSKONQILRIENHD 289  
DB 240 GMPAADAGAGAGNLDPLFRNSQFOALRAMVQANPQILQPMVLVELSKONQILRIENHD 299  
QY 290 EFLQLINPEPEGSGDPLDQPEDEMPHAISVTPPEQEAIGRLSMGDFRARIAPLAC 349  
DB 300 DFLRLINEPVE-GEENVMEQLEA-AMPOAVTTPPEEREALERLEGMGDFRAMVLEVPFAC 357  
QY 350 DRNEELAANYLLEHAGE-ED 368  
DB 358 NKNEELAANYLLDHMEFED 377

RESULT 9  
Q94C35 PRELIMINARY; PRT; 378 AA.  
ID Q94C35  
AC Q94C35;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical 40.1 kDa protein.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Shinn P.,  
RA Tracy S.E., Banh J., Bowser L., Carninci P., Chung M.K.,  
RA Goldsmith A.D., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,  
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Pham P.K., Quach H.L.,  
RA Sakano H., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C.,  
RA Toriumi M., Yamada K., Yu G., Shinozaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT "Arabidopsis cDNA clones";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AY037181; AAKS9766.1;  
DR InterPro; IPR004806; Rad23.  
DR InterPro; IPR006636; ST11.  
DR InterPro; IPR000449; UBA domain.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00627; UBA; 2.  
DR Pfam; PF00240; ubiquitin; 1.  
DR SMART; SM00727; ST11; 1.  
DR SMART; SM00165; UBA; 2.  
DR SMART; SM00213; UBQ; 1.

DR TIGRFAMs; TIGR00601; rad23; 1.  
DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 378 AA; 40075 MW; E49B591527A3EC27 CRC64;

Query Match 48.7%; Score 930.5; DB 10; Length 378;  
Best Local Similarity 52.6%; Pred. No. 4.6e-58;  
Matches 200; Conservative 56; Mismatches 109; Indels 15; Gaps 8;

QY 1 MKLTVKTLKGTGHEIRVQPNDDTMAVKKNIEEIOGKDSYPMGQQLIFNGKVLKDESTLE 60  
DB 1 MKLIFVKTLSSGNEFEIEVKPADKVDKTAETVKGAE-YPAAKQMLIHQGVKLDSTLE 59  
QY 61 ENKVNEDGFLVVMLSKGKTS---GSTGTSSSHSNTPATRQAPPLEAP-QQAPQPPVAPI 116  
DB 60 ENNVVENSFVIMLSKTKASPSGASTASAPASATQPTVATPQVSAPTASVPVTSGTA 119  
QY 117 TTSQPEGLPAQAPNTHDAAASNLSSGRNVDTIINQLMGGGSGWDKQVORALRAAYNNP 176  
DB 120 TAAAPATAASVQTDVYQGAASNLVAGTTLESTVQQLIDMGSGSWDRDVTVVRALRAAFNNP 179  
QY 177 ERAVEYLISGIPVTAETI-----AVPIGGQGAN--TTDRAPTGEAGLSGIPNTAPLDLFPQ 229  
DB 180 ERAVEYLISGIPVTAETI-----AVPIGGQGAN--TTDRAPTGEAGLSGIPNTAPLDLFPQ 239  
QY 230 GASNAGGAGGGLDPLFRNNPQFOAVREVMVHTNPQILQPMVLVELSKONQILRIENHD 289  
DB 240 GMPAADAGAGAGNLDPLFRNSQFOALRAMVQANPQILQPMVLVELSKONQILRIENHD 299  
QY 290 EFLQLINPEPEGSGDPLDQPEDEMPHAISVTPPEQEAIGRLSMGDFRARIAPLAC 349  
DB 300 DFLRLINEPVE-GEENVMEQLEA-AMPOAVTTPPEEREALERLEGMGDFRAMVLEVPFAC 357  
QY 350 DRNEELAANYLLEHAGE-ED 368  
DB 358 NKNEELAANYLLDHMEFED 377

RESULT 10  
Q40742 PRELIMINARY; PRT; 392 AA.  
ID Q40742  
AC Q40742;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE OsrAD23.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nipponbare;  
RA MEDLINE=97369378; PubMed=9225866;  
RX Schultz T.F., Quatrano R.S.;  
RT "Characterization and expression of a rice RAD23 gene.";  
RL plant Mol. Biol. 34:557-562(1997).  
DR EMBL; U63530; AAB65841.1;  
DR HSSP; P54725; LDV0.  
DR Gramene; Q40742; -.  
DR InterPro; IPR004806; Rad23.  
DR InterPro; IPR006636; ST11.  
DR InterPro; IPR000449; UBA domain.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00627; UBA; 2.  
DR Pfam; PF00240; ubiquitin; 1.  
DR SMART; SM00727; ST11; 1.  
DR SMART; SM00165; UBA; 2.  
DR SMART; SM00213; UBQ; 1.  
DR TIGRFAMs; TIGR00601; rad23; 1.  
DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
SQ SEQUENCE 392 AA; 41754 MW; BD6E08574CC7C6B CRC64;



Query Match 46.5%; Score 887.5; DB 10; Length 392;  
 Best Local Similarity 50.9%; Pred. No. 5.4e-55;  
 Matches 199; Conservative 52; Mismatches 109; Indels 31; Gaps 10;

QY 1 MKLTKVTLKGTTHPEIRVQPNNDTIMAVKNIIEEIQGKDSYPWGQQLLIIFNGKVLKDESTLE 60  
 DB 1 MKISVTKLKGSTFQIEVDSQAQVADVKRIETTCQGHYPAEQQMLIHQGVKLVKDDTTLD 60

QY 61 ENKVNEDGFLVWMLSGTSGTSSQHSNTPATQAPLEAPQAP-OPPVPITTT 118  
 DB 61 ENKVLNEDGFLVWMLSGTSGTSSQHSNTPATQAPLEAPQAP-OPPVPITTT 116

QY 119 -----SQEGLPAQAP-----NTHDNAASLLSGRNVDTIINOLMENGSGSWDKD 163  
 DB 117 VPTVSAPTTATAPAPAVASVSADNVGQATSNLVAGSNLEATIQLSEMGGGIWD 176

QY 164 KVORALRAAYNPERAVEYLISGIPVTAIEIAPVIGG-QCANTTDRA-----PTGEAGLSGI 218  
 DB 177 IVLHALSAFNNPERAVEYLSGVPEQMDIPVPPPSIQPANPTQASQATQAPPAAPSISSG 236

QY 219 PNTAPLDFPQOASNAGGAGG-GPLDFLRNNPQAVREMYHTNPNQILQMLVLSKON 277  
 DB 237 PNASPLDFPQALPNASTDAAGLNLDARNNAQRTLLSLVQANPQILQPLQLGLKON 296

QY 278 PQILRLIBENHDFQLLNEPFG-GEQDFLDQ-PEEDMPHRAISVTPEEQEAGRLISM 335  
 DB 297 PQILQLIQENQAEFLHLINEPAEGDDEENLLDQFPE--AMPQTIAVTFEDEAILRLEPM 354

QY 336 GFDRARVTEAFIACDRNBEALANYLLEHAGE 366  
 DB 355 GFDRALVLDVFACNKDBQLAANYLLDHME 385

RESULT 11

Q9S9L8  
 ID Q9S9L8 PRELIMINARY; PRT; 246 AA.  
 AC Q9S9L8  
 DT 01-MAY-2000 (TremBLrel. 13, Created)  
 DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
 DE T24D18.27 protein.  
 GN T24D18.27.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Liu S.X., Yu G., Sakano H., Jhaveri A., Lee J.M., Lenz C., Pham P.,  
 RA Toriumi M., Chin C., Chlou J., Choi E., Chung M., Gonzalez A.,  
 RA Hwang B., Koo T., Li J., Liu A., Vayberg M., Altafi H., Brooks S.,  
 RA Buehler E., Chao Q., Conn L., Conway A., Hansen N., Johnson-Hopson C.,  
 RA Khan S., Kim C., Lam B., Nguyen M., Palm C., Shinn P., Tambunga G.,  
 RA Davis R.W., Ecker J.R., Federpiehl N.A., Theologis A.;  
 RT "The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Theologis A.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC010924; AAF18513.1; -.  
 DR HSSP; P02248; 1UBI.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 1.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR SMART; SM00165; UBA; 1.  
 DR SMART; SM00213; UBA; 1.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.

SQ SEQUENCE 246 AA; 25966 MW; 0867D1E7D72B6FFE CRC64;  
 Query Match 34.6%; Score 661.5; DB 10; Length 246;  
 Best Local Similarity 56.7%; Pred. No. 3.2e-39;  
 Matches 143; Conservative 29; Mismatches 63; Indels 17; Gaps 4;

QY 1 MKLTKVTLKGTTHPEIRVQPNNDTIMAVKNIIEEIQGKDSYPWGQQLLIIFNGKVLKDESTLE 60  
 DB 1 MKLTKVTLKGTTHPEIRVQPNNDTIMAVKNIIEEIQGKDSYPWGQQLLIIFNGKVLKDESTLE 60

QY 61 ENKVNEDGFLVWMLSGTSGTSSQHSNTPATQAPLEAPQAP-OPPVPITTT 120  
 DB 61 ENKVTGEGFLVWMLSGTSGTSSQHSNTPATQAPLEAPQAP-OPPVPITTT 108

QY 121 PEGLPAPAPNTHDNAASLLSGRNVDTIINOLMENGSGSWDKQVQALRAAYNPERAV 180  
 DB 109 VQEQPTAQSDTYGQAATSLVSGSIEQVQIMEMGGSGDKETVTTRALRAAYNPERAV 168

QY 181 EYLSGIPVTAIEI-APVIGGQANTTDRAPTGEAGLSGIPNTAPLDFPQOASNAGGAG 239  
 DB 169 DYLSGIPETVTPATNLGSGVSGRELTAAPP-----SGGNSSPLDLFPQEAVIDAAGD 224

QY 240 GGPLDFLRNNPQ 251  
 DB 225 LGTLEFLRGNDQ 236

RESULT 12

Q8WUB0  
 ID Q8WUB0 PRELIMINARY; PRT; 409 AA.  
 AC Q8WUB0  
 DT 01-MAR-2002 (TremBLrel. 20, Created)  
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
 DE RAD23 homolog B (S. cerevisiae).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC020973; AAH20973.1; -.  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR006636; STI1.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR SMART; SM00727; STI1; 1.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; UBO; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 SQ SEQUENCE 409 AA; 43199 MW; 475PBD499DACAC69 CRC64;

Query Match 32.5%; Score 621.5; DB 4; Length 409;  
 Best Local Similarity 34.6%; Pred. No. 4.5e-36;  
 Matches 146; Conservative 89; Mismatches 156; Indels 67; Gaps 11;

QY 1 MKLTKVTLKGTTHPEIRVQPNNDTIMAVKNIIEEIQGKDSYPWGQQLLIIFNGKVLKDESTLE 60  
 DB 1 MQVTLKTLQOOTFKIDIDPEETVKALKKEIESEKGDAPFVAGQKLIYAGKILNDTALK 60

QY 61 ENKVNEDGFLVWMLSGTSGTSSQHSNTPATQAPLEAPQAP-OPPVPITTT 106  
 DB 61 EYKIDKQNFVVMVKPAVSTAPATQASAPASTTAVTSSTTTTVAQPTPPALAPT 120

QY 107 QAP-----QF-PVAPITTSQEPGLPAQAP-----NTHDNA 135  
 DB 121 STPASITPASATASSEPAPASAAKQEPKPAETPVATSTATDSTSGDSSRSNLPEDA 180



QY 136 ASNLSSGRNVDITINQLMEMGGSDKQVQALRAAYNNPRAVEYLYSGIPVTAI-A 194  
 DB 181 TSALVTQSYENWVTEINSMG---VEREQVIAALRASNNPRAVEYLYLGMIPGDRBSQA 237  
 QY 195 VPIGGCANTTDRAPTEAGLSGIPNTAPLDLFPQASNAGGAGGGLDFLRNNPQQA 254  
 DB 238 VVDPFQAAT--GVPOSSAVAAATTT-----ATTTSSTSGHPLFLRNQPFQ 287  
 QY 255 VREMVHTNPQILQPLMLVELSKNPQILRIENHDEFLQLNEPFE-----GGEGDFLD 308  
 DB 288 MQIIQONSLPALLOQIGRENPLQLOISQHEFIQMLNEPVEAGCGGGGGSG 347  
 QY 309 QPEDEMPHA--ISVTPQEQAIGRLSMGFRARVIEAFLACDRNEELAANYLLEHAGE 366  
 DB 348 GIAEAGSGHMYIQTVPQEKAEIRKALGFPEGLVIQAFACKENLAANFLQQNFD 407  
 QY 367 ED 368  
 DB 408 ED 409

RESULT 13  
 Q8CAP3 PRELIMINARY; PRT; 362 AA.  
 AC Q8CAP3;  
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE RAD23a homolog.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium.  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK038300; BAC29962.1;  
 SQ SEQUENCE 362 AA; 39635 MW; 804B2608ECA241A9 CRC64;

Query Match 31.4%; Score 599; DB 11; Length 362;  
 Best Local Similarity 33.7%; Pred. No. 1.5e-34;  
 Matches 134; Conservative 84; Mismatches 108; Indels 72; Gaps 9;

QY 3 LTVTKLGTHTFEIRVQPNDTIMAVKNIETIQGKDSYPMWQQLIFNGKVLKDESTLEEN 62  
 DB 5 ITLTKLQQQTFTEIRVQPNDTIMAVKNIETIQGKDSYPMWQQLIFNGKVLKDESTLEEN 64  
 QY 63 KVNEDGFLVNLKSGTSGTGTSSQHSNTPATROAPPLEAPQ-QAPQPPV----- 113  
 DB 65 HIDEKNFVWVTKAKAG-----QGIPAPPEASPTAVPSPFPFVLASGMSHP 114  
 QY 114 -----APITTSQPLPAQAPNT-----HDNAASNLSSGRNVDITINQLMEMG 156  
 DB 115 PPTSREDSKPSSESTTTTSPESIGSVSSGSSGREDAASTLVTSGEYETMLTEINSMG 174  
 QY 157 GGSWDKQVQALRAAYNNPRAVEYLYSGIPVTAIETAVPIGGCANTTDRAPTEAGLS 216  
 DB 175 ---YERERVAALRASNNPRAVEYLYLTGIP-----GSPPEHSGVQESAP----- 219  
 QY 217 GIPNTAPLDLFPQASNAGGAGGGLDFLRNNPQQAVERMVHTNPQILQPLMLVELSKQ 276  
 DB 220 -----EQPATE---AGENFLEFLRDQFQFQNMQRQVIOQNPAALLPALLQQLGQE 264  
 QY 277 NPQILRIENHDEFLQLNEP-----FEGEGDFLDQPEDEMPHAISVTPPEQAIG 330  
 DB 265 NPQLLQOISRQEQFIQMLNEPPEGLADISDVEGEVGAIGEAPQMNYIQTTPQKEAIE 324

QY 331 RLESMGFRARVIEAFLACDRNEELAANYLLEHAGEBD 368  
 DB 325 RLKALGPESLVIOYAFACKENLAANFLLSQNFDE 362

RESULT 14  
 Q9XZE0 PRELIMINARY; PRT; 414 AA.  
 ID Q9XZE0  
 AC Q9XZE0;  
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)  
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE DR23 OR DHR23 OR CG1836.  
 GN Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nabirochikina E.N., Grischuk A.L., Soldatov A.V.;  
 RT "Cloning and characterization of the Drosophila melanogaster homologue  
 RT of the Sacharomyces cerevisiae gene RAD23."  
 RL Genetika 35:0-0(1999).  
 DR EMBL; AF136606; AAD33695.1; --  
 DR HSSP; P54725; IDV0.  
 DR FlyBase; FBgn0026777; Rad23.  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR006636; STI1.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR SMART; SM00727; STI1; 1.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS50053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 414 AA; 45794 MW; 4C2FF94C116F7AB CRC64;

Query Match 25.7%; Score 490.5; DB 5; Length 414;  
 Best Local Similarity 29.0%; Pred. No. 9.2e-27;  
 Matches 129; Conservative 77; Mismatches 130; Indels 109; Gaps 11;

QY 1 MKLTVTKLGTHTFEIRVQPNDTIMAVKNIETIQGKDSYPMWQQLIFNGKVLKDESTLE 60  
 DB 1 MIITIKLQQQTFTEIRVQPNDTIMAVKNIETIQGKDSYPMWQQLIFNGKVLKDESTLE 59  
 QY 61 ENKVNEDGFLVNLKSGTSGTGTSSQHSN-----TPATROAPPLEAPQAPQP----- 111  
 DB 60 SYNVDKKFIVMLTRDSSSNRNQLSVKESNKLSTDDSKQSMPCERANHTNSPSTNT 119  
 QY 112 -----PVAPITTSQPLPAQAPNTHDNAASNLSSGRNVDITINQLMEMGGSDKQK 164  
 DB 120 EDSVLSRETRPUSDDSLIGELAQ--SLQSRASNLMDGEYNTVLSWEMG---YPREQ 175  
 QY 165 VQALRAAYNNPRAVEYLYSGIPV-----TAEIAPVPIGGCANTTDRAPTE 212  
 DB 176 VERAMPAASNNPRAVEYLYINGIPAEEGTFYRNELNSTNPISLPISGFPASATSAERSTE 235  
 QY 213 AGLSGIPNTAPLDLFPQASNAGGAGGGLDFLRNNPQQAVERMVHTNPQILQPLMLVE 272  
 DB 236 -----SNSDPFELRSQPPQLMRSLIYQNPHLLHAVLQQ 270  
 QY 273 LSKONPQILRIENHDEFLQLNEP-----EGG----- 302  
 DB 271 IGOTNPALLQILSENQDAFLMLNQLPDRSESGATVPVSNARIPTLDNVDLFPDLE 330  
 QY 303 -----EGDFLDQPEDEMPHAIS---VTPPEQAIGRLSMGFRARV 342  
 DB 331 VATSAQRSAGTSAHSGSAAADN-EDLEQPLGVSTIRLNQDKDAIERLKALGFPEALV 389

```

Qy 343 IEAFLACDRNEELAANYLLEHAGEE 367
Db 390 LQAYFACEKNEEQAAFLSSSFED 414

RESULT 15
Q9V3W9 PRELIMINARY; PRT; 414 AA.
AC Q9V3W9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE RAD23 protein.
GN RAD23 OR CG1836.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Brodsky M.H., Rubin G.M., Tsang G.;
RT "Full Length Drosophila melanogaster cDNA sequence.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003844; AAF59352.1; --
DR EMBL; AF132147; AAD33594.1; --
DR HSSP; P54725; IDV0.
DR FlyBase; FBgn0026777; Rad23.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STII.

```

```

DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR000626; UbiQuitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiQuitin; 1.
DR SMART; SM00727; STII; 1.
DR SMART; SM00165; UBA; 1.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS0053; UBIQUITIN 2; 1.
SQ SEQUENCE 414 AA; 45780 MW; 4C2E494CA116F7AB CRC64;

Query Match 25.5%; Score 487.5; DB 5; Length 414;
Best Local Similarity 28.8%; Pred. No. 1.5e-26;
Matches 128; Conservative 78; Mismatches 130; Indels 109; Gaps 11;

Qy 1 MKLTIVKTLKGTHTFEIRVQPNDDTMAVKNTIEETQKDSYFWGQQLLIFNGKVLKDESTLE 60
Db 1 MIITIKNLQOQTFTIEFAPEKTVLELKKKIFEEKGPE-YVAEKOKLIYAGVILTDRTVG 59
Qy 61 ENKVNEDGFLVWMLSKGKTSGTSTSSOHSN-----TPATRQAPPLEAPQQAPOP----- 111
Db 60 SYNVDKKFIVMLTRDSSSNRNQLSVKESNKLKLTSTDDSKQKMPCEANHTNPSSTNT 119
Qy 112 -----PVAPITTSQPEGLPAQAPNTHDNAASNLSSGRNVDTIINOLMENGSGGSKDK 164
Db 120 EDSVLSRETRPLSSDELIGELAQA-SLQSRASENLLMGDEYNQTVLSWVENG-----YPREQ 175
Qy 165 VQALRAAYNPERAVEYLYSGIPV-----TABIAPVIGGGANTTDRAPTGE 212
Db 176 VERAMAASYNPERAVEYLYNGIPAEKGTFNRLNENSTSLIPSGQPASATSAERSTE 235
Qy 213 AGLSGIENAPLDLPQGNASGAGGGLDFLNNPQFQAVREMYHTNPQILQPLVLE 272
Db 236 -----SNSDPFEFLRSQFQLMKRSLIYQNPHELLHAVLQQ 270
Qy 273 LSKQNPOILRLIENHDEFLQLNEPF-----EGG----- 302
Db 271 IGQTNPALQLISENQDAFLNMLNQPIDRESESGATVPVSNARIPSTLDNVDLFSFDLE 330
Qy 303 -----EGDFLQPEDEMPHAIS-----VTPEQEAIGRLMESMGFDRAV 342
Db 331 VATSAQRSAAGTSAHQSGAADN-EDLEQPLGVSTIRLNQKDAIERLKALGFPEALV 389
Qy 343 IEAFLACDRNEELAANYLLEHAGEE 367
Db 390 LQAYFACEKNEEQAAFLSSSFDD 414

```

Search completed: December 17, 2003, 06:20:07  
Job time : 38.6093 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:16:39 ; Search time 35.229 Seconds  
(without alignments)  
1658.049 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 1910

Sequence: 1 MKLTVKTLKGTFFIRVQPN.....CDRBEAANYLLEHAGEED 368

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1910	100.0	368	21 AAY71459	Maize Rad23 protei
2	1169.5	61.2	371	21 AAG17377	Arabidopsis thalia
3	1115	58.4	368	21 AAG44341	Arabidopsis thalia
4	1067.5	55.9	348	21 AAG17378	Arabidopsis thalia
5	1014	53.1	345	21 AAG44342	Arabidopsis thalia
6	931.5	48.8	378	21 AAG19974	Arabidopsis thalia
7	931.5	48.8	378	21 AAG45204	Arabidopsis thalia
8	900.5	47.1	405	21 AAY71458	Maize Rad23 protei
9	871.5	45.6	299	21 AAG17379	Arabidopsis thalia

10	834	43.7	335	21	AAG19975	Arabidopsis thalia
11	834	43.7	335	21	AAG45205	Arabidopsis thalia
12	820	42.9	296	21	AAG44343	Arabidopsis thalia
13	742	38.8	307	21	AAG19976	Arabidopsis thalia
14	742	38.8	307	21	AAG45206	Arabidopsis thalia
15	634	33.2	416	23	ABBS7171	Mouse ischaemic co
16	625.5	32.7	409	24	ABU07460	Protein differenti
17	622.5	32.6	409	19	AAW75700	Vpr protein bindin
18	622.5	32.6	409	19	AAW68186	Vpr binding protei
19	604	31.6	117	23	ABP34618	Human ORF3591 prot
20	603.5	31.6	363	18	AAW23658	EGAP-binding prote
21	603.5	31.6	363	19	AAW75699	Vpr protein bindin
22	603.5	31.6	363	19	AAW68185	Vpr binding protei
23	603.5	31.6	375	24	ABU11879	Human ABCA1 intera
24	603	31.6	379	21	AAW58841	Breast and ovarian
25	574.5	30.1	346	18	AAW21730	Nuclear mitotic ap
26	563.5	29.5	257	21	AAG36529	Arabidopsis thalia
27	487.5	25.5	414	22	ABBS5854	Drosophila melanog
28	466	24.4	214	21	AAG36530	Arabidopsis thalia
29	375	19.6	290	22	ABB72011	Drosophila melanog
30	374	19.6	186	21	AAG36531	Arabidopsis thalia
31	356	18.6	179	19	AAW75843	Vpr protein bindin
32	356	18.6	179	19	AAW68187	Vpr binding protei
33	333	17.4	174	19	AAW75850	Protein sequence o
34	333	17.4	174	19	AAW68194	Protein sequence o
35	308.5	16.2	136	21	AAG33942	Arabidopsis thalia
36	292.5	15.3	150	19	AAW75849	Protein sequence o
37	292.5	15.3	150	19	AAW68193	Protein sequence o
38	217	11.4	112	19	AAW75848	Protein sequence o
39	217	11.4	112	19	AAW68192	Protein sequence o
40	195.5	10.2	536	21	AAG30040	Arabidopsis thalia
41	195.5	10.2	577	21	AAG30039	Arabidopsis thalia
42	182	9.5	79	20	AAW87979	Ubiquitin-like dom
43	177	9.3	589	22	AAW94311	Human protein sequ
44	176	9.2	464	21	AAW12447	Human 149Y2H151 u
45	176	9.2	589	22	ABUS2717	Human brain-derive

#### ALIGNMENTS

##### RESULT 1

AAW71459

ID AAY71459 standard; Protein; 368 AA.

XX

AC AAY71459;

XX

DT 04-OCT-2000 (first entry)

XX

DE Maize Rad23 protein #2.

XX

XX

KW Rad23; maize; ATCC No: PTA-531; recombinant expression cassette; wheat;

KW transgenic plant; soybean; sunflower; sorghum; canola; modulator.

XX

OS Zea mays.

XX

PN WO200031268-A1.

XX

PD 02-JUN-2000.

XX

PF 12-OCT-1999; 99WO-US24129.

XX

PR 23-NOV-1998; 98US-0109728.

XX

PA (PION-) PIONEER HI-BRED INT INC.

XX

PI Mahajan PB, Tagliani L;

XX

DR WPI; 2000-400078/34.

DR N-PSDB; AAD01231.

XX

PT Isolated nucleic acid encoding maize RAD23 protein is used to modulate

the levels of polypeptides in plant or in assays for identifying

PT compounds that bind to and/or increase/decrease enzymatic activity of  
XX catalytically active polypeptides -  
XX Claim 11b; Page 78-79; 82pp; English.  
CC The present sequence is the maize Rad23 protein #2. It is isolated from  
CC a Zea mays cell line, B73 callus tissue regenerated five days after  
CC transfer of the callus from medium containing auxin to a medium devoid  
CC of exogenous auxin. The cDNA is deposited under the ARCC No: PTA-531.  
CC Maize Rad23 DNA sequence operably linked to a promoter, can be used to  
CC construct a recombinant expression cassette. This expression cassette  
CC can be used to generate a dicot or monocot transgenic plant e.g., maize,  
CC soybean, sunflower, sorghum, canola, wheat, etc. It can also be used to  
CC modulate the levels of Rad23 polypeptide expression in a plant or in  
CC assays to identify compounds, that bind to and/or modulate the enzymatic  
CC activity of catalytically active polypeptides.  
XX  
SQ Sequence 368 AA;  
  
Query Match 100.0%; Score 1910; DB 21; Length 368;  
Best Local Similarity 100.0%; Pred. No. 7.9e-147;  
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MKLTIVTKLKGTHFEIRVOPNDTIMAVKKNIEIQKDSYPWGQQLLIENGKVLKDESTLE 60  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1 MKLTIVTKLKGTHFEIRVOPNDTIMAVKKNIEIQKDSYPWGQQLLIENGKVLKDESTLE 60  
QY 61 ENKVNEDGLVVLMSKGTSGTGTSSQHSNTPATROAPLEAPQAPQPPVAPITTSQ 120  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
61 ENKVNEDGLVVLMSKGTSGTGTSSQHSNTPATROAPLEAPQAPQPPVAPITTSQ 120  
QY 121 PEGLPAPQPNTHDAAASLLSGRNVDTIINQLMENGSGSWDKVQRAIRAAYNPERAV 180  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
121 PEGLPAPQPNTHDAAASLLSGRNVDTIINQLMENGSGSWDKVQRAIRAAYNPERAV 180  
QY 181 EYLYSGIPVTAEIPIGQGGANTTDRAPTGAGISGIPNTAPLDLFPQASNAGGAGG 240  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
181 EYLYSGIPVTAEIPIGQGGANTTDRAPTGAGISGIPNTAPLDLFPQASNAGGAGG 240  
QY 241 GPLDLFRNNPQAVREVMVHTNPQILOPMLVELSKONPQILRLIEENHDEFLLQLNPEP 300  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
241 GPLDLFRNNPQAVREVMVHTNPQILOPMLVELSKONPQILRLIEENHDEFLLQLNPEP 300  
QY 301 GEGDFLDQDEDEPHAISVTPEQAEIGRLESNGFDRARVIEAFLACDRNEELAANYL 360  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
301 GEGDFLDQDEDEPHAISVTPEQAEIGRLESNGFDRARVIEAFLACDRNEELAANYL 360  
QY 361 LEHAGEED 368  
DB |||||||  
361 LEHAGEED 368  
  
RESULT 2  
AAG17377  
ID AAG17377 standard; Protein; 371 AA.  
XX  
AC AAG17377;  
XX  
XX 17-OCT-2000 (first entry)  
DT  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 18372.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 03-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132048.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.

```
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0147260.
PR 27-JUL-1999; 99US-0147303.
PR 27-JUL-1999; 99US-0147319.
PR 28-JUL-1999; 99US-0148341.
PR 02-AUG-1999; 99US-0147493.
PR 02-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 61.2%; Score 1169.5; DB 21; Length 371;
Best Local Similarity 64.3%; Pred. No. 1.2e-86;
Matches 247; Conservative 43; Mismatches 65; Indels 29; Gaps 9;

QY 1 MKLTVKTLKTHFEIRVQPNDTIMAVKKNIEIRIQKDSYPMGQQLIFNGKVLKDESTLE 60
DB 1 MKLTVKTLKSHFEIRVLPFSDTIMAVKKNIEDSQKNDYPCQQLIHNGKVLKDETSLV 60
QY 61 ENKVNEDGLVVLMSKGTSGTSSSOHSN-----TPATQAPPLEAPQAPQP 112
DB 61 ENKVTBEGFLVVLMSKSGSAGQSVQTSVSPVSAATTSSTKPAAP--STTQSSPVP 118
QY 113 VAPITTSQPEGLPAQAPNTHDNAENLLSGRNVDTIINQMEMGGGWDKVKQVRAALAA 172
DB 119 ASPIPAQEQ---PAAQTDTYGAATLYSGSSLEQMVQOIMEMGGGWDKVTTRALAA 175
QY 173 YNNPERAVLYSGIPVTAETAVP-----IGGGANTTTDRAPTGBAGLSGIPNTAPLDLF 227
DB 176 YNNPERAVDLYSGIPQTAEVAVPPEAQIAGSGA-----ADVAPA--SGGPNSSPLDLF 228
QY 228 POGASNAGGAGGGLDFLNNPQFAVREWVHTNPQILQPMVLKSKNPQILRIEEN 287
DB 229 PQETVAAGSGDLGTLEFLRNNDQFQQLRTWVHSNPQILQPMVLKSKNPQILRIEEN 288
QY 288 HDEFLLQLNEPPEG--GEGDFLDQPEDEMPHAI SVTPEEQAI GRLESMPDRARVIEA 345
DB 289 QAEFLQLVNEPYEGSDGEGDMFQPEQ-EMPHAINVTAEQAIQRLEANGFDRLVIEA 347
QY 346 FLACDRNBELAANYLLEHAGE-ED 368
DB 348 FLACDRNBELAANYLLENSGDPED 371

RESULT 3
AAG44341
ID AAG44341 standard; Protein; 368 AA.
XX
```

AC AAG44341;  
XX 18-OCT-2000 (first entry)  
DT  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55531.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139464.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.

```
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0160770.
PR 18-OCT-1999; 99US-0160814.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-016192.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 58.4%; Score 1115; DB 21; Length 368;
Best Local Similarity 62.8%; Pred. No. 3,26-82;
Matches 235; Conservative 45; Mismatches 82; Indels 12; Gaps 8;

Qy 1 M K L T V K T L G T H F I R V Q P N D T I N A V K K N I E E I Q G D S Y P W G Q C L L I F N G K V L K D E S T L E 60
Db 1 M K L T V K T L G S H F I R V L P T D T I N A V K K N I E D S Q S K D N Y P G Q Q L L I H N G K V L K D E T T L V 60

Qy 61 E N K Y N D G F L V M L S K G T S G T S S S Q - H S N T P A T R Q A P P L E A P Q A P P V P A P I T T S 119
Db 61 E N K Y T E G F L V V M L S K S T A S S A P S T Q P T I T T S I S T T L A A P - S T T Q S T A V P A S N S 119

Qy 120 Q P - E G L P A Q P A N T H N A S N L L S G R N V D T I N Q L M E M G G S W D K K V O R A L R A A Y N N P R 178
Db 120 T P V Q E Q T A Q S D T Y G Q A A S T L V S G S S I E Q M V Q Q I M E M G G S W D K E T V T R A L R A A Y N N P R 179

Qy 179 A V E Y L Y S G I P T A E I - A V P I G G Q G A N T T D R A P T G E A G L S G I P N T A P L D L F P Q A S N A G G 237
Db 180 A V D Y L Y S G I P T V T I P A T N L S G V G S G R E L T A P P P - - - - S G P N S P L D L F P Q E A V S D A A G 235
```

```
Qy 238 A C G G P L D F L R N N P Q F O A V R E M V H T N P Q I L Q P M L V E L S K Q N P Q I L R L I E N H D B F L Q L L N E 297
Db 236 G D L G T L E F L R G N D F O O L R S M V N S P Q I L Q P M L Q E L G K Q N P Q L R L I Q E N Q A S F L Q L L N E 295

Qy 298 P F E G G E G - D P L D P E D E M P H A I S V T P R E O E A I G R L E S M G F P R A R V I E A F L A C D R N E E L 355
Db 296 P Y E G S D G V D I F O P D Q - E M P H S V N V T P E Q E S I E R L E A M G F D R A I V I E A F L S C D R N E E L 354

Qy 356 A A N Y L L E H A G E - E D 368
Db 355 A A N Y L L E H S A D F E D 368

RESULT 4
AAG17378
ID AAG17378 standard; Protein; 348 AA.
XX
AC AAG17378;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 18373.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 23-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
```

PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139819.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144634.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 16-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 55.9%; Score 1067.5; DB 21; Length 348;  
Best Local Similarity 62.9%; Pred No. 2,1e-78;  
Matches 227; Conservative 41; Mismatches 64; Indels 29; Gaps 9;



QY	24	MAVKKNIEIQKDSYPMQQLIFNGKVLKDESTLEENKKNEDGFLVVMLSKGTSST	83
Db	1	MAVKKNIEDSQKNDYPCQQLIHNGKVLKDETSILVENKVTEEGFLVVMLSKSGGSA	60
QY	84	GTSSQSHN-----TPATQAPPLEAPQAQPPVAPITTSQPEGLUPAQAPNTHNA	135
Db	61	GOASVQTSSVSQFVSATTSSTKPAAP--STTQSSPVPASPIPAQBQ---PAAQTDTYQQA	115
QY	136	ASNLLSGRNVDTIINOLMEMGGGSKDKVKQALRAAYNNPERAVEYLYSGIPVTAETAV	195
Db	116	ASTLVSGSLGQVQOIMEMGGGSKDKETVTRALRAAYNNPERAVDLYSGIPQTAETAV	175
QY	196	P-----IGCGGANTTDRAPTGAGLSGIPENTAPLDLPQGNAGGAGGGPLDLFRNPP	250
Db	176	PVPEAQIAGSGA----APVAPA--SGGPNSSPLDLFPQETVAAGSGDLGLTLELRND	228
QY	251	QQAQVREMYHTNPQILOPMLVELSKONPOILRLIENHDEFLOLLNEPPEG--GSGDFLD	308
Db	229	QFQQLRTMYSNPQILOPMLQBLGKQPOLLRLIQENQAFTQLVNEPYEGSDGSGDMFD	288
QY	309	QPEEDEMPHAISVTPPEQEAIGRLSMGPDRAVIEAFACDRNBELAANYLLEHAGE-E	367
Db	289	QPEQ-EMPHAINVTPAEQAIORLEAMGPDRAVIEAFACDRNBELAANYLLENSGDFE	347
QY	368	D 368	
Db	348	D 348	
RESULT 5			
AAG44342			
ID	AAG44342 standard; Protein; 345 AA.		
AC	AAG44342;		
DT	18-OCT-2000 (first entry)		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 55532.		
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
OS	Arabidopsis thaliana.		
PN	EP1033405-A2.		
PD	06-SEP-2000.		
PF	25-FEB-2000; 2000EP-0301439.		
XX	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	28-APR-1999;	99US-0130891.	
PR	30-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	04-MAY-1999;	99US-0132407.	
PR	05-MAY-1999;	99US-0132484.	
PR	06-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	01-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	16-JUN-1999;	99US-0139453.	
PR	17-JUN-1999;	99US-0139492.	
PR	18-JUN-1999;	99US-0139454.	
PR	18-JUN-1999;	99US-0139455.	
PR	18-JUN-1999;	99US-0139456.	
PR	18-JUN-1999;	99US-0139457.	
PR	18-JUN-1999;	99US-0139463.	
PR	18-JUN-1999;	99US-0139750.	
PR	18-JUN-1999;	99US-0139763.	
PR	21-JUN-1999;	99US-0139817.	
PR	22-JUN-1999;	99US-0139899.	
PR	23-JUN-1999;	99US-0140353.	
PR	23-JUN-1999;	99US-0140354.	
PR	24-JUN-1999;	99US-0140695.	
PR	28-JUN-1999;	99US-0140823.	
PR	29-JUN-1999;	99US-0140991.	
PR	30-JUN-1999;	99US-0141287.	
PR	01-JUL-1999;	99US-0141842.	
PR	01-JUL-1999;	99US-0142154.	
PR	02-JUL-1999;	99US-0142055.	
PR	06-JUL-1999;	99US-0142390.	
PR	08-JUL-1999;	99US-0142803.	
PR	09-JUL-1999;	99US-0142920.	
PR	12-JUL-1999;	99US-0142977.	
PR	13-JUL-1999;	99US-0143542.	
PR	14-JUL-1999;	99US-0143624.	
PR	15-JUL-1999;	99US-0144005.	
PR	16-JUL-1999;	99US-0144085.	
PR	16-JUL-1999;	99US-0144086.	
PR	19-JUL-1999;	99US-0144325.	
PR	19-JUL-1999;	99US-0144331.	
PR	19-JUL-1999;	99US-0144332.	
PR	19-JUL-1999;	99US-0144333.	
PR	19-JUL-1999;	99US-0144334.	
PR	19-JUL-1999;	99US-0144335.	
PR	20-JUL-1999;	99US-0144352.	
PR	20-JUL-1999;	99US-0144632.	
PR	20-JUL-1999;	99US-0144884.	
PR	21-JUL-1999;	99US-0144814.	
PR	21-JUL-1999;	99US-0145086.	
PR	21-JUL-1999;	99US-0145088.	
PR	22-JUL-1999;	99US-0145087.	
PR	22-JUL-1999;	99US-0145089.	
PR	22-JUL-1999;	99US-0145192.	

Wed Dec 17 06:49:41 2003

PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145318.  
PR 27-JUL-1999; 99US-0145319.  
PR 27-JUL-1999; 99US-0145391.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159299.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160982.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161362.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161992.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162143.  
PR 29-OCT-1999; 99US-0162144.

Query Match 53.1%; Score 1014; DB 21; Length 345;  
Best Local Similarity 61.3%; Pred. No. 4.7e-74;  
Matches 215; Conservative 44; Mismatches 80; Indels 12; Gaps 8;

QY 24 MAVKKNIEIQKDSYPMGQQLLIIFNGKVLKDESTLEENKVNEDGLVVMLSKGTSGST 83  
DB 1 MAVKKNIEDSQSKDNYPCGQQLLIHNGKVLKDETLVENKVTETGFLVVMLSKSTASSA 60  
QY 84 GTSSSQ-HSNTPATRQAPPLEAPQAPQPPVAPITTSQP-EGLPAPQAPNTHDNASNLIS 141  
DB 61 GPSSTQPTSTTSTISSITTLAAP-STTQSIAPASNSTPVOPTAQSDTYGQAATLVS 119  
QY 142 GRNVDTIINQMEMGGGSHDKVQALRAAAYNNPERAVEYLYSGIPVTAET-AVPIGQ 200  
DB 120 GSSIEQMVQIIMEMGGGSHDKETVTRALRAAYNNPERAVDYLYSGIPETVTPATNL 179  
QY 201 GANTTDRAPTGEAGLSGIPTAPLDLFPQASNAGGAGGGLDFLRNNPQFOAVREMH 260  
DB 180 GSGRELTAPPP-----SGGNSPFLDFPQENVSDAAGGLTLEFLRGNDQFOQLRSMVN 235  
QY 261 TNPQILQPMVLVELSKONQILRLIENHDEFLLQNLNPPFEGGEG--DFLDQPEDEMPHA 318  
DB 236 SNFQILQPMQLQELGKONQILRLIQENQAEFLQLNPEYEGSDGVDIFDQPDQ-EMPHS 294  
QY 319 ISVTPEQEAIGRLSMGDFDRARVIEAFIACDNEELAANYLLEHAGE-ED 368  
DB 295 VNVTPQEESIERLEAMGDFRAIVIEAFILSCDRNEELAANYLLEHAGEAD 345

RESULT 6  
AAG19974  
ID AAG19974 standard; Protein; 378 AA.  
XX AAG19974;  
AC AAG19974;  
XX 17-OCT-2000 (first entry)  
DT 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21981.  
XX Arabidopsis thaliana.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
PD 25-FEB-2000; 2000EP-0301439.  
PF 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134318.  
PR 14-MAY-1999; 99US-0134319.  
PR 14-MAY-1999; 99US-0134321.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135829.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142877.  
PR 13-JUL-1999; 99US-0143342.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.



```
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0143920.
PR 12-JUL-1999; 99US-0143977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 48.8%; Score 931.5; DB 21; Length 378;
Best Local Similarity 52.6%; Pred No. 2.7e-67;
Matches 200; Conservative 56; Mismatches 109; Indels 15; Gaps 8;

Qy 1 MKLTVKTLKGTHTFEIRYQPNDDTINAVKKNIEBIOGKDSYPMGQQLIFNGKVLKDESTLE 60
Db 1 MKIFVKTLGSGNFEIEVKPADKVDVKTATETVKGAB-YPAAKQMLIHQGVKLDETTLE 59

Qy 61 ENKYNEDGFLVWMLSKGKTS---GSTGTSSQHSNTPATQAPPLEAP-QOAPOPPVAPI 116
Db 60 ENNVENSFVIMLSKTKASPGASTASAPASATQPTVATPQVSAPTASVPVPTSGTA 119

Qy 117 TTISOEGLPAQAPNTHDNAASNLGSRNVDTIINQLEMWGGGWDKQVORALRAAYNNP 176
Db 120 TAAAPATAASVQTDVYCGAASNLVAGTTLSTVQOILDMGGGWDRTVVRALRAAFNNP 179

Qy 177 ERAVEYLYSGIPVTAETI-----AVPIGGQGAN--TTDRAPTGEAGLSGIPNTAPLDLFPQ 229
Db 180 ERAVEYLYSGIPVTAETI-----AVPIGGQGAN--TTDRAPTGEAGLSGIPNTAPLDLFPQ 239

Qy 230 GASNAGGAGGGPLDFLRNPPQAVREMHVNTNPILOPMLVELSKONPQILRLIENHND 289
Db 240 GMPAADAGAGAGNLDFLRNSQQFOALRAMVQANPQILQPMQLQELGKQNPQLVRLIQEHOA 299

Qy 290 EFLQLNPEPGEGBDFLDQPEEDEMHPAISVTPPEOEATGRLESMDGFRARVIEAFAC 349
Db 300 DFLRLINEPVE-GEENVMEQLEA-AMPQAVTVTPPEEREATERLEGMGFDRAMVLEVPFAC 357

Qy 350 DRNEELAANYLLEHAGE-ED 368
```



PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149446.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match  
Best Local Similarity

45.6%; Score 871.5; DB 21; Length 299;  
60.3%; Pred. No. 1.5e-62;

Matches		188;	Conservative	36;	Mismatches	59;	Indels	29;	Gaps	9;
QY	73	MLSKGKTSGTSSQHSN-----TPATRQAPPEAPQAPQPPVAPITTSQPEGL	124							
Db	1	MLSKSKSGGAGQASVQTSSVPSATTSSTKPAAP--STTQSSVPVAPSPAPQSQ---	55							
QY	125	PAQAPNTHNAANLISGRNVDTIINQLMEMGGGSKDKVQBALRAAYNNPERAVEYLX	184							
Db	56	PAAQTDTYGAASLTGSGSLEQMVQOIMEMGGGSKDKETVTRALRAAYNNPERAVDYLY	115							
QY	185	SGIPVTAEIAVP-----IGGOGANTTDRAPTGEAGLSGIPNTAPLDLFPQAGNAGGAG	239							
Db	116	SGIPQTAEVAVPVEPAQIAGSGA-----APVAP--SGGNSPDLDFQETVAAAGSGD	168							
QY	240	GGPLDFLRNNPQFQAVREMYHTNPQILQPLMLVELSKONPQILRLIENHDEFQLLNPEP	299							
Db	169	LGTLEFLRNDQFQQLRTMVHSNPQILQPLMLQELGKQPQLLIQENQAEFLQLVNEPY	228							
QY	300	EG--GEGDFLDQPEDEMPHAI SVTPEQEAIGRLSEMGDFRVRVTEAFACDRNEELAA	357							
Db	229	EGSDGEGDMFQPEQ-EMPHAINVTFAEQEAIQRLEAMGDFRALVTEAFACDRNEELAA	287							
QY	358	NYLLEHAGE-ED 368								
Db	288	NYLLESGDFED 299								
RESULT 10										
AAG19975										
ID	AAG19975 standard; Protein: 335 AA.									
XX										
AC	AAG19975;									
XX										
DT	17-OCT-2000 (first entry)									
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 21982.									
XX										
KW	Protein identification; signal transduction pathway; metabolic pathway;									
KW	hybridisation assay; genetic mapping; gene expression control; promoter;									
KW	termination sequence.									
XX										
OS	Arabidopsis thaliana.									
XX										
FN	EP1033405-A2.									
XX										
PD	06-SEP-2000.									
XX										
PF	25-FEB-2000; 2000EP-0301439.									
XX										
PR	25-FEB-1999;	99US-0121825.								
PR	05-MAR-1999;	99US-0123180.								
PR	09-MAR-1999;	99US-0123548.								
PR	23-MAR-1999;	99US-0125788.								
PR	25-MAR-1999;	99US-0126264.								
PR	29-MAR-1999;	99US-0126785.								
PR	01-APR-1999;	99US-0127462.								
PR	06-APR-1999;	99US-0128234.								
PR	08-APR-1999;	99US-0128714.								
PR	16-APR-1999;	99US-0129845.								
PR	19-APR-1999;	99US-0130077.								
PR	21-APR-1999;	99US-0130449.								
PR	23-APR-1999;	99US-0130510.								
PR	28-APR-1999;	99US-0130891.								
PR	30-APR-1999;	99US-0132048.								
PR	04-MAY-1999;	99US-0132407.								
PR	05-MAY-1999;	99US-0132484.								
PR	06-MAY-1999;	99US-0132485.								
PR	06-MAY-1999;	99US-0132486.								
PR	06-MAY-1999;	99US-0132487.								
PR	07-MAY-1999;	99US-0132863.								
PR	11-MAY-1999;	99US-0134256.								
PR	14-MAY-1999;	99US-0134218.								

PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUN-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144632.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.



PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 43.7%; Score 834; DB 21; Length 335;  
Best Local Similarity 53.3%; Pred. No. 1.9e-59;  
Matches 179; Conservative 47; Mismatches 96; Indels 14; Gaps 7;  
  
Qy 45 LLIFNGKVLDESTEENKVNEDFLVVLKSGKTS---GSTGTSSSOHSNTPATROAPP 101  
Db 1 MLIHGKVLKDETELENNVVENSVFVIMLSKTKASPSGASTASAPASATOPQVATPQ 60  
  
Qy 102 LEAP-QQAPQPPVAPITTSQPEGLPAQAPNTHDAAASLLGRNVDTTIINLMWGGGWSW 160  
Db 61 VSAPTASVPVPTSGTATAAATAASVQTDVYQASNLVAGTTLSTVQVILDMGGGWSW 120  
  
Qy 161 DKQVQRALRAAYNNPERAVEYLYSGIPVTABI-----AVPIGGQGAN--TTDRAPTGEA 213  
Db 121 DRDVTVRALRAAFNNPERAVEYLYSGIPAQAABI PPVAQAPATGEGQAANPLAQPOQAAAPA 180  
  
Qy 214 GLSGIPNTAPLDLPQGNASNAGGGGGLDPLRNPNQFOAVRENVHTNPQILOMIVEL 273  
Db 181 AATGGPNANPLNLPQGMPPAADAGAGNLDPLRNSQQFOALRAMVQANPQILOPMLQEL 240  
  
Qy 274 SKQNPQLRLIEENHDEFLQLNPPFEGGEGDLDQPEDEMPHAISVTPPEQEAIGRLE 333  
Db 241 GKQNPQLVRLIOEHQADFLRLINEPVE-GEENVMEQLEA-AMPQAVTVTPPEREAIERLE 298  
  
Qy 334 SMGFDRARVIEAFLACDRNEELAANYLLEHAGE-ED 368  
Db 299 GMGFDRAVLEVFACNKEELAANYLLDHMHPEP 334  
  
RESULT 11  
AAG45205  
ID AAG45205 standard; Protein; 335 AA.  
AC AAG45205;  
XX  
XX 18-OCT-2000 (first entry)  
DT  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56724.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 20-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142800.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.

[illegible]

## RESULT 12

AAG44343

ID AAG44343 standard: Protein: 296 AA

XX  
ID  
AM947343 STANDARD; FIOLETH; 250 AA.

AC AAG44343:

AC  
XX  
AHG44343;

DT 18-OCT-2000 (first entry)

18-OCT-2000 (first entry)  
xx

XX DE VERBODENDE HET VERBODENDE VERBODENDE

DE  
yy  
Arabidopsis thaliana protein fragment

**EXHIBIT**

protein identification; signal transduction

hybridisation assay; genetic mapping

KW  
5012  
termination sequence.

XXV

OS *Arabidopsis thaliana*.

XX

PN EP1033405-A2.

XX  
-----

PD 06-SEP-2000.

XX

PF	25-FEB-1999	2000EP-03014339
XX	25-FEB-1999	99US-0121825
PR	09-MAR-1999	99US-0123180
PR	05-MAR-1999	99US-0123548
PR	23-MAR-1999	99US-0125788
PR	25-MAR-1999	99US-0126785
PR	29-MAR-1999	99US-0127462
PR	01-APR-1999	99US-0128234
PR	06-APR-1999	99US-0128714
PR	08-APR-1999	99US-0129845
PR	16-APR-1999	99US-0130077
PR	19-APR-1999	99US-0130449
PR	21-APR-1999	99US-0130510
PR	23-APR-1999	99US-0131419
PR	23-APR-1999	99US-0130891
PR	28-APR-1999	99US-0131449
PR	30-APR-1999	99US-0132048
PR	04-MAY-1999	99US-0132407
PR	04-MAY-1999	99US-0132484
PR	05-MAY-1999	99US-0132485
PR	06-MAY-1999	99US-0132486
PR	06-MAY-1999	99US-0132487
PR	11-MAY-1999	99US-0132863
PR	17-MAY-1999	99US-0134256
PR	14-MAY-1999	99US-0134218
PR	14-MAY-1999	99US-0134219
PR	14-MAY-1999	99US-0134221
PR	14-MAY-1999	99US-0134370
PR	18-MAY-1999	99US-0134768
PR	19-MAY-1999	99US-0134941
PR	20-MAY-1999	99US-0135124
PR	21-MAY-1999	99US-0135353
PR	24-MAY-1999	99US-0135629
PR	25-MAY-1999	99US-0136021
PR	27-MAY-1999	99US-0136392
PR	28-MAY-1999	99US-0136782
PR	01-JUN-1999	99US-0137222
PR	03-JUN-1999	99US-0137528
PR	04-JUN-1999	99US-0137502
PR	07-JUN-1999	99US-0137724
PR	08-JUN-1999	99US-0138094
PR	10-JUN-1999	99US-0138540
PR	10-JUN-1999	99US-0138847
PR	14-JUN-1999	99US-0139119
PR	16-JUN-1999	99US-0139452
PR	16-JUN-1999	99US-0139453
PR	17-JUN-1999	99US-0139459
PR	18-JUN-1999	99US-0139454
PR	18-JUN-1999	99US-0139455
PR	18-JUN-1999	99US-0139456
PR	18-JUN-1999	99US-0139457
PR	18-JUN-1999	99US-0139458
PR	18-JUN-1999	99US-0139459
PR	18-JUN-1999	99US-0139460
PR	18-JUN-1999	99US-0139461
PR	18-JUN-1999	99US-0139462
PR	18-JUN-1999	99US-0139463
PR	18-JUN-1999	99US-0139750
PR	21-JUN-1999	99US-0139763
PR	21-JUN-1999	99US-0139817
PR	22-JUN-1999	99US-0139899
PR	23-JUN-1999	99US-0140353
PR	23-JUN-1999	99US-0140354
PR	24-JUN-1999	99US-0140695
PR	28-JUN-1999	99US-0140823
PR	30-JUN-1999	99US-0140991
PR	30-JUN-1999	99US-0141287
PR	01-JUL-1999	99US-0141842
PR	02-JUL-1999	99US-0142154
PR	06-JUL-1999	99US-0142390
PR	08-JUL-1999	99US-0142380

PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	13-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 42.9%; Score 820; DB 21; Length 296;		
Best Local Similarity 58.3%; Pred. No. 2.2e-58;		
Matches 176; Conservative 40; Mismatches 74; Indels 12; Gaps 8;		
QY	73	MLSKGKTSSTGSSQ-HSNTPATRQAPPLPAPQAPPPVAPITTSOP-EGLPQAQPN 130
Db	1	MLSKSKTASSAGPSSTQPTSTTSSITTLAAP-STTQSIAPASNSTPVQRPQTAQSD 59
QY	131	THDNAASNLGRNVDTIINQLMEMGGGSKDKVORALRAAYNNPERAVEVLYSGIPVT 190
Db	60	TYGQAASLTLYSGSSIEQMVGQIEMEMGGGSKDKETVTRALRAAYNNPERAVDYISGIPET 119
QY	191	AEI-AVPIGGOGANTTDRAPTGEAGLSGIPNTAPLDLPQAGSNAGGAGGGLDFLRNN 249
Db	120	VTIPATNLSGVSGRELTAAPP-SSGGNSSPLDLFPQEAVSDAAGGDLGTLEFLRGN 175
QY	250	PQFQAVREVMVHTNPQILQPMVLVSKQNQILRIENHDEFLQNLNEPPEGEG--DFL 307
Db	176	DQFQQLRSMVSNPQILQPMVLQELKQNQILRIQENQAEFLQNLNEPTEGSDGVDIF 235
QY	308	DQPEEDEMFAISVTPPEQEAIGRLSMGDFDRARVEAFACDRNEELAANYLLEHAGE- 366
Db	236	DQPDQ-EMPHSVNVTPPEQESIERLEAMGDFRAIVIEAFLSCDRNEELAANYLLEHSAF 294
QY	367	ED 368
Db	295	ED 296
RESULT 13		
AAG19976		
ID	AAG19976 standard; Protein; 307 AA.	
XX		
AC	AAG19976;	
XX		
DT	17-OCT-2000 (first entry)	



PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 38.8%; Score 742; DB 21; Length 307;		
Best Local Similarity 52.3%; Pred. No. 5.1e-52;		
Matches 161; Conservative 42; Mismatches 91; Indels 14; Gaps 7;		
Qy	73	MLSKGKTS---GSGTSSSHSNTPATRQAPPLEAP-QQAPQPPVAPITTSQEGUQAQ 128
Db	1	MLSKTKASPGSASTAPASATOPQTATPQVSAPTASVEPTSGTATAAPATAASVQ 60
Qy	129	PNTHDNASLLSRNVDTIINQMENGSGSWDKVQRALRAAYNNPERAVEVLYSGIP 188
Db	61	TDVYQAAASNLVAGTTLETSTVQQLDMMGGGSDRTTVRALRAAFNNPERAVEVLYSGIP 120
Qy	189	VTAEI-----AVPTGGGAGN--TTDRAPTGEAGISGIPNTAPLDLFPQGNAGGAGGG 241
Db	121	AQAEIPVQAQATGEQANPLAQQAAPAAATGFPNANPLNLFQGMFADAGAGAG 180
Qy	242	PLDFLRNNPQAVREVMHTNPQILQPMVLVELSKONFQILRLIEENHDFLQLNEPFG 301
Db	181	NLDFLRNSQQQALRAVQANPQILQPMVLQELKONFQVRLIQHQADFLRLINEPVE- 239
Qy	302	GEGDFLOQPEDEPHALISVTPPEQEAIGRLSGFDRARVIEAFACDRNEELAANYLL 361
Db	240	GEENVMEQLEA-AMFQAVTVTPPEAREATERLEGMGDFDRAMVLEVFACNKEELAANYLL 298

Qy	362	EHAGE-ED 368
Db	299	DHMEPED 306
RESULT 14		
AAG45206		
ID	AAG45206 standard; Protein; 307 AA.	
XX	AAG45206;	
DT	18-OCT-2000 (first entry)	
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 56725.	
DE	Arabidopsis thaliana	
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
OS	Arabidopsis thaliana.	
XX	EP1033405-A2.	
PD	06-SEP-2000.	
XX	25-FEB-2000; 2000EP-0301439.	
XX	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.



Db 61 TDVYGOASNLVAGTTLESTVQOILDMGGSDWDRTVVRALRAAFNNPRAVEYLYSGIP 120  
QY 189 VTAEI-----AVPIGQGAN--TTDRAPTGEAGLSGIPNTAPLDLFPQASNAGGAGGG 241  
Db 121 AQAEIPVQAAPATGEOAANPLAQPOOAAAAPAAATGGPNANPLNLPFGMPAADAGAGAG 180  
QY 242 PLDFLENNPQFQAVREWHNTNPOILQPMVLVELSKNPQILRLTEENHDSFLOLLNEPPG 301  
Db 181 NLDFLENSOQFQALRAMVQANPQILQPMQLQEGKQNPQLVRLIQHQADFRLINEPVB- 239  
QY 302 GEGDFLDQPEEDEMHPAISVTPEQEAIGRLSMGFDRAVIEAFIACDRNEELAAANYLL 361  
Db 240 GEENWEQLEA-AMPQAVTTPEREAIERLEGMGFDRAWLVEVFFACNKEELAAANYLL 298  
QY 362 EHAGE-ED 368  
Db 299 DHMHEPED 306

RESULT 15  
ABB57171  
ID ABB57171 standard; Protein; 416 AA.  
XX  
AC ABB57171;  
XX  
DT 07-MAR-2002 (first entry)  
XX  
DE Mouse ischaemic condition related protein sequence SEQ ID NO:415.  
XX  
KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
KW vasospastic ischaemia; ischaemic condition; ischaemic disease.  
XX  
OS Mus musculus.  
XX  
XX WO200188188-A2.  
XX  
XX 22-NOV-2001.  
XX  
XX 18-MAY-2001; 2001WO-JP04192.  
XX  
XX 18-MAY-2000; 2000JP-0145977.  
XX  
XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
XX  
XX PA  
XX PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
XX  
XX WPI; 2002-034733/04.  
XX  
XX N-PSDB; ABI99464.

Examining the ischemic condition (e.g. occlusive ischemia) by measuring expression levels of particular genes defined in the specification or by determining the expression profile of a gene group comprising these genes -

Claim 2; Page 1132-1134; 2690pp; English.  
The present invention describes a method for examining ischaemic conditions, comprising measuring the expression levels of particular genes (I) in a test sample or determining the expression profile of a gene group in the sample comprising genes selected from (I). The method is useful for examining the ischaemic condition (e.g. compressive ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring expression levels of particular genes (ABI99202 to ABI99912, encoding the protein sequences in ABB57020 to ABB57374) or by determining the expression profile of a gene group comprising these genes. The expression levels or expression profiles produced by these genes are used as an indicator when screening for ischaemic condition-improving drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR primers for a mouse ischaemic condition related sequence, which are used in the exemplification of the present invention.

XX Sequence 416 AA;

Query Match 33.2%; Score 634; DB 23; Length 416;  
Best Local Similarity 33.5%; Pred. No. 4.6e-43;  
Matches 146; Conservative 93; Mismatches 109; Indels 88; Gaps 11;  
QY 1 MKLVTTLKGTHTFEIRVOPNDTIMAVKNIEIOCKDSYPWGQQLLIIFNGKVLKDESTLE 60  
Db 1 MCVTLKTLQQOQTFKIDIDPBEETVKALKKEIKSEKGDAPPVAGQKLIYAGKILSDDTALK 60  
QY 61 ENKVNEDFLVVMLSKGTSGTGTSSSHQHNTPA---TROAPPLEAPQ-OAQPQPPVAPI 116  
Db 61 EYKIDKFNVVVMVTKRAVTTAVPATTPQSTPSTPTTVSSSPAVAAAQAAPTALAPT 120  
QY 117 TT-----SQPE---GLPAQAP-----NTHDNA 135  
Db 121 STPATSTTASSEPAPAGATQPKPAEKPAQTFVLTSAPADSTPCDSSRSNLFEDA 180  
QY 136 ASNLLSGRNVDTIINQLMEMGGGSHDKVORALRAAYNNPERAVEYLYSGIPVTAETAV 195  
Db 181 TSALVTGQSYENMVTEIMSMG---YEREQVTAALRAFNFNDRAVEYLLMGIPGDRSQQA 237  
QY 196 PIGGOGANTTDRAPTGEAGLSGIPNTAPLDLFPQASNAG-----GGAGGGPLDFLRN 248  
Db 238 VVD-----PPQAVSTGTPOS-----PAVAAAATTTATTTTSGHPLFLRN 281  
QY 249 NPQFOAVREWHNTNPOILQPMVLVELSKNPQILRLIEENHDEFLLNEPFE-----300  
Db 282 QPQFOQMRQIIQONPSLLPALLOQIGRENPNLLOQISQHEHFIQMLNEPVOEAGGQGGG 341  
QY 301 -----GGEGDFLDQPEEDEMHPAISVTPEQEAIGRLSMGFDRAVIEAFIACDRN 352  
Db 342 GGGGGGGGGGGGGGIAEAGSGHMNY-IQVTPQEKAEIRLKGALGPFPEGLVIOAYFACEKN 400  
QY 353 BELAANYLLEHAGEED 368  
Db 401 ENLAANFLQQNFDED 416

Search completed: December 17, 2003, 06:18:39  
Job time : 36.229 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:20:11 ; Search time 126.158 Seconds  
(without alignments)  
544.771 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 1910

Sequence: 1 MKLTVKTLKGTHTFIRVQPN.....CDRNEELANYLLEHAGEED 368

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pdp:\*  
12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pdp:\*  
13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1910	100.0	368	9	US-09-805-550-4
2	900.5	47.1	405	9	US-09-805-550-2
3	603.5	31.6	363	12	US-10-116-275-170
4	603	31.6	379	15	US-10-102-806-549
5	182	9.5	79	11	US-09-918-036-4
6	176	9.2	589	12	US-10-293-000-4
7	176	9.2	595	12	US-10-293-000-2
8	164	8.6	81	11	US-09-918-036-5
9	154	8.1	522	15	US-10-296-770-2
10	150	7.9	624	15	US-10-146-473-47
11	140	7.3	504	11	US-09-918-864-410
12	140	7.3	504	12	US-10-195-835-410
13	140	7.3	504	12	US-10-286-333-410
14	140	7.3	504	15	US-10-125-635A-410
15	140	7.3	504	15	US-10-002-603-410

16	139	7.3	63	9	US-09-925-299-1544	Sequence 1544, Ap
17	139	7.3	63	11	US-09-925-299-1544	Sequence 1544, Ap
18	134	7.0	421	15	US-10-296-770-7	Sequence 7, Appli
19	133.5	7.0	1274	15	US-10-020-215-2	Sequence 2, Appli
20	131	6.9	139	12	US-10-263-828-120	Sequence 120, App
21	130	6.8	966	12	US-10-304-454-2	Sequence 2, Appli
22	129	6.8	655	14	US-10-001-632A-2	Sequence 2, Appli
23	129	6.8	655	15	US-10-054-683-29	Sequence 29, Appl
24	128.5	6.7	336	10	US-09-987-107-44	Sequence 44, Appl
25	128.5	6.7	390	15	US-10-102-806-461	Sequence 461, App
26	128	6.7	337	10	US-09-987-107-46	Sequence 46, Appl
27	127	6.6	388	12	US-10-094-749-2859	Sequence 2859, Ap
28	126.5	6.6	160	10	US-09-738-973-205	Sequence 205, App
29	126.5	6.6	160	10	US-09-854-133-205	Sequence 205, App
30	126.5	6.6	160	15	US-10-144-649A-205	Sequence 205, App
31	126.5	6.6	228	12	US-10-029-386-33570	Sequence 33570, A
32	124.5	6.5	710	15	US-10-296-770-4	Sequence 4, Appli
33	123.5	6.5	110	9	US-09-925-299-1042	Sequence 1042, Ap
34	123.5	6.5	110	11	US-09-925-299-1042	Sequence 1042, Ap
35	123	6.4	151	12	US-10-360-053-10	Sequence 10, Appl
36	122.5	6.4	76	9	US-09-804-866-16	Sequence 16, Appl
37	122.5	6.4	76	9	US-09-826-312-13	Sequence 13, Appl
38	122.5	6.4	76	11	US-09-918-036-1	Sequence 1, Appli
39	122.5	6.4	76	14	US-10-058-820-21	Sequence 21, Appl
40	122.5	6.4	76	15	US-10-108-767-13	Sequence 13, Appl
41	122.5	6.4	76	15	US-10-152-156-13	Sequence 13, Appl
42	122.5	6.4	76	15	US-10-275-985-1	Sequence 1, Appli
43	122.5	6.4	128	12	US-10-205-194-87	Sequence 87, Appl
44	122.5	6.4	159	12	US-10-360-053-12	Sequence 12, Appl
45	122.5	6.4	530	15	US-10-296-770-5	Sequence 5, Appli

#### ALIGNMENTS

#### RESULT 1

US-09-805-550-4  
; Sequence 4, Application US/09805550  
; Patent No. US20020026045A1  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; TITLE OF INVENTION: Rad23 Genes and Uses Thereof  
; FILE REFERENCE: 0964D  
; CURRENT APPLICATION NUMBER: US/09/805,550  
; CURRENT FILING DATE: 2001-03-13  
; PRIOR FILING DATE: 09/413,574  
; PRIOR FILING DATE: 1999-10-06  
; PRIOR APPLICATION NUMBER: 60/109,728  
; PRIOR FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-805-550-4

Query Match	100.0%	Score	1910;	DB	9;	Length	368;
Best Local Similarity	100.0%;	Pred. No.	5.6e-148;				
Matches	368;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MKLTVKTLKGTHTFIRVQPN	DTIMAVKKNIEETQKDSYPMGQOLIFNKGVLKDESTLE	60			
Db	1	MKLTVKTLKGTHTFIRVQPN	DTIMAVKKNIEETQKDSYPMGQOLIFNKGVLKDESTLE	60			
Qy	61	ENKVNEDGFLVWMLSKGTSGTGTSSQHSNTPATRQAPPLAPQAPPPVAPITTSQ	120				
Db	61	ENKVNEDGFLVWMLSKGTSGTGTSSQHSNTPATRQAPPLAPQAPPPVAPITTSQ	120				
Qy	121	PEGLPAQAPNTHDAAANLLSGRNVDTIINQLMEMGGSGWDDKQVQALRAAYNNPERAV	180				
Db	121	PEGLPAQAPNTHDAAANLLSGRNVDTIINQLMEMGGSGWDDKQVQALRAAYNNPERAV	180				

QY 181 EYLYSGIPVTABIAVPIGGQANTTDRAPTBAGLSGIPNTAPLDLFPQASNAGGGAGG 240  
Db 181 EYLYSGIPVTABIAVPIGGQANTTDRAPTBAGLSGIPNTAPLDLFPQASNAGGGAGG 240  
QY 241 GPLDFLRNNPQAVREVMHTNPQILOPMLVLSKONPQILRLIENHDEFLOLLNEPFE 300  
Db 241 GPLDFLRNNPQAVREVMHTNPQILOPMLVLSKONPQILRLIENHDEFLOLLNEPFE 300  
QY 301 GEGDFLDQPEDEMPHATSVTPEQEAIGRLSMGDFRARIYEAFLACDRNEELAANYL 360  
Db 301 GEGDFLDQPEDEMPHATSVTPEQEAIGRLSMGDFRARIYEAFLACDRNEELAANYL 360  
QY 361 LEHAGEED 368  
Db 361 LEHAGEED 368

RESULT 2  
US-09-805-550-2  
; Sequence 2, Application US/09805550  
; Patent No. US2002026045A1  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; TITLE OF INVENTION: Rad23 Genes and Uses Thereof  
; FILE REFERENCE: 0964D  
; CURRENT APPLICATION NUMBER: US/09/805,550  
; CURRENT FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 09/413,574  
; PRIOR FILING DATE: 1999-10-06  
; PRIOR APPLICATION NUMBER: 60/109,728  
; PRIOR FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 405  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-805-550-2

Query Match 47.1%; Score 900.5; DB 9; Length 405;  
Best Local Similarity 49.0%; Pred. No. 2.4e-65;  
Matches 197; Conservative 53; Mismatches 117; Indels 35; Gaps 7;  
QY 1 MCLTVKTLKGTHTPEIRVQPNDDTMAVKNIIEIQQKDSYPWGOQLLIFNGKVLKDESTLE 60  
Db 1 MCLNVKTLKGTHTPEIRVQPNDDTMAVKNIIEIQQKDSYPWGOQLLIFNGKVLKDESTLE 60  
QY 61 ENKYNEDGFLVWMLSKGTSSTSSQHSNTPATRQAP-----PLEAPQAP-QPPVA 114  
Db 61 SNGVAENSLVIMLSKAKAS--SSGASTATTAKAPATLAQPAAPVAPAAVARPTQAPVA 119  
QY 115 PITTSQPEGLPAQAP-----NTHDNAASLLSGRNVDTIINQMGGSGWDKDK 164  
Db 120 TAETAPSVQQAAPATAATVAATDADVYQAASNLVFGNNLEQTIQIILDMGGGTWERDT 179  
QY 165 VQALRAAYNNPERAVEYLSGIPVTAE-----IAVPIGGQANTTDRAPTGEA----- 213  
Db 180 VVRLAAYNNPERAIDYLSGIPENVEAQPVAPAAAGQOONQOAAAPQAPVALVPQ 239  
QY 214 -GLSGIPNTAPLDLFPQASNAGG-----GAGGGLDFLRNNPQFQAVREVMHTNPQIL 266  
Db 240 SPASAGPNANLNLFPQVPGSGNPGVVPAGSGALDALRLQFQALLQVLQANPQIL 299  
QY 267 QPMLVLSKONPQILRLIENHDEFLOLLNEPFEFGEGDFLDQPEDEMPHATSVTPEEQ 326  
Db 300 QPMLQELKQNPQILRLIQENQAEBFLRVNESPESGPGCNILGQLAAAVPQTLVTPTEER 359  
QY 327 BAIGRLSMGDFRARIYEAFLACDRNEELAANYLLEHAGEED 368  
Db 360 BAIGRLSMGDFRARIYEAFLACDRNEELAANYLLEHAGEED 401

RESULT 3  
US-10-116-275-170  
; Sequence 170, Application US/10116275  
; Publication No. US20030211476A1  
; GENERAL INFORMATION:  
; APPLICANT: Elan Pharmaceutical Technology  
; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Brayden, David  
; APPLICANT: Byrne, Daragh  
; APPLICANT: Lambkin, Imelda  
; APPLICANT: Higgins, Lisa  
; TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and Methods and  
; FILE REFERENCE: E1067/20087  
; CURRENT APPLICATION NUMBER: US/10/116,275  
; CURRENT FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 349  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 170  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-116-275-170

Query Match 31.6%; Score 603.5; DB 12; Length 363;  
Best Local Similarity 33.4%; Pred. No. 4e-41;  
Matches 133; Conservative 88; Mismatches 106; Indels 71; Gaps 8;  
QY 3 LTVKTLKGTHTPEIRVQPNDDTMAVKNIIEIQQKDSYPWGOQLLIFNGKVLKDESTLEEN 62  
Db 5 I7LKTQQQTFAIRMEPDETIVKLEKIEAEKGRDAFPVAGQKLIYAGKILSDDDVPIRDY 64  
QY 63 KYNEDGFLVWMLSKGTSSTSSQHSNTPATRQAPLEAPQAPQPPVPAPIT----- 117  
Db 65 RIDENKFNVMVMT--KTRAGQGT-----APPEASPTAPESSTPPPPAPTSGMSHP 114  
QY 118 -----TSQPEGLPAQAPNT-----HDNAASLLSGRNVDTIINQMGG 156  
Db 115 PPAAREDKSPSEASPTTSPESVSGSVSSGSGREDAASTLVGTSEYETMLTEIMSG 174  
QY 157 GSGWKKDKVORALRAAYNNPERAVEYLSGIPVTABIAVPIGGQANTTDRAPTGEAGLS 216  
Db 175 ---YERERVVAALRASYNPNPRAVYLLTGIP-----GSPPEHSGVQESQVS 219  
QY 217 GIPNTAPLDLFPQASNAGGAGGGLDFLRNNPQFQAVREVMHTNPQILOPMLVLSKQ 276  
Db 220 EQPATE-----AAGENPLEFLRQPPQFQNNRQVITQNNPALLPALLQQLGOE 265  
QY 277 NQOILRLIENHDEFLOLLNEP-----FEGEGDFLDQPEDEMPHATSVTPEEQBAIG 330  
Db 266 NPQLLOQISRHOEQFIQMLNEPPGELADISDVEGEVGAIGEAPQWNYIQTVPQKEAIE 325  
QY 331 RLESMDGDFRARIYEAFLACDRNEELAANYLLEHAGEED 368  
Db 326 RLKALGFPSELVIOATFACEKNLANFLLSQNFDE 363

RESULT 4  
US-10-102-806-549  
; Sequence 549, Application US/10102806  
; Publication No. US20030054421A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA103P1C1  
; CURRENT APPLICATION NUMBER: US/10/102,806  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/925,298  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05881  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270



```

; FILE REFERENCE: 4115-175
; CURRENT APPLICATION NUMBER: US/10/293,000
; CURRENT FILING DATE: 2003-04-01
; PRIOR APPLICATION NUMBER: US 60/338,549
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-000-2

Query Match      9.2%; Score 176; DB 12; Length 595;
Best Local Similarity 21.2%; Pred. No. 7.4e-06;
Matches 99; Conservative 56; Mismatches 143; Indels 168; Gaps 19;

Qy 1 MKLVTKLGTHTFEIRVQPNDDTMAVKNIETEEQKDSYPMWQQLIFNGKVLKDESTLE 60
Db 37 MKVTVKTPKEKE-EFAVPENSSVQOFK---EEISKRFKSHTDQLVLIFAGKILKDDQTLS 92

Qy 61 ENKVNEDGFLVVMLSK-----GKTSGST----- 83
Db 93 QHGIIH-DGLTVHLVIKTRNPQDHSAAQNTAGTNVTTSTPNSNSTSGSATSNPFLGLG 151

Qy 84 -----GTSSQHSNTPATROAPPLEAPQAPQ----- 110
Db 152 LGGIAGLSLGLNTNFSLSQSMQRLSNPEMMVMQIMENPFVQSMLSNPDLMRQLIMA 211

Qy 111 -PPVAPITTSQPE-----GLPAQAPNTHDNAASLLSGRNVDTTIINQLMWSGGSD 161
Db 212 NPQMQQLIQNPISHLNPNPDINRQTLERAPNMMQEMRNQDRALSLEIPGG--- 268

Qy 162 KDKVQRALRAAYNNPER-----AVELYSGIPVTAEIAVPIGGQ-----ANTTDRAP----- 209
Db 269 -----YNALRMVTDIQEPMLSAAQEQGGNPFASLVNSTSSGEGSQPSRTENRDLNPNW 324

Qy 210 -----TGE--AGLSGIPNTAPDLFPQASNAGGAGGGPDLFL 246
Db 325 APQTSQSSASSGSTASTVGGTGTSTAGTSQSTTAP-NLVP-----GVGAS-----M 371

Qy 247 RNNPQFAVRMVTNPQILQPMI-----VELSKNPQILRLIBENHDEFQLINPEP 298
Db 372 FNTQMOSLLOQITENPQLMQMLLSAPYMRSMQSLSQNPDLAAQM-----MLNPP 422

Qy 299 FEGEGED-----FLDQPEDEMPHAISVTPPEQEAIGRL 333
Db 423 LFAGNPQLQEQMRQLPTFLQMQNPDTLSAMS-NPRAMQALLQIQ 467

RESULT 8
US-09-918-036-5
; Sequence 5, Application US/09918036
; Publication No. US2003092159A1
; GENERAL INFORMATION:
; APPLICANT: MADURA, Kiran
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE RAPID PURIFICATION OF PROTEASOME
; TITLE OF INVENTION: METHODS OF USE OF COMPONENTS THEREOF
; FILE REFERENCE: 266/165
; CURRENT APPLICATION NUMBER: US/09/918,036
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/050,171
; PRIOR FILING DATE: 1997-06-19
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-918-036-5

Query Match      8.6%; Score 164; DB 11; Length 81;
Best Local Similarity 38.2%; Pred. No. 4.4e-06;

Matches 29; Conservative 26; Mismatches 21; Indels 0; Gaps 0;

Qy 3 LTVTKLGTHTFEIRVQPNDDTMAVKNIETEEQKDSYPMWQQLIFNGKVLKDESTLE 62
Db 5 ITLKLTQQTQKIRMEPDETIVKVLKEIEAEKGRDAFPVAGOKLIYAGKILSDDDVPDIRY 64

Qy 63 KVNEDGFLVVMLSK 78
Db 65 RIDKKNFVVVMVTKTK 80

RESULT 9
US-10-296-770-2
; Sequence 2, Application US/10296770
; Publication No. US20030104570A1
; GENERAL INFORMATION:
; APPLICANT: Caberson Silva, Teresa Elisa Virginia
; APPLICANT: Delisse, Anne-Marie Eva Bernande
; TITLE OF INVENTION: Triple Fusion Proteins Comprising
; TITLE OF INVENTION: Ubiquitin Fused Between Thioredoxin and a Polypeptide of
; TITLE OF INVENTION: Interest
; FILE REFERENCE: B45221
; CURRENT APPLICATION NUMBER: US/10/296,770
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: PCT/EP01/06952
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: GB 0015619.0
; PRIOR FILING DATE: 2000-06-26
; PRIOR APPLICATION NUMBER: GB 0026484.6
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Chimaeric (E. coli - human)
US-10-296-770-2

Query Match      8.1%; Score 154; DB 15; Length 522;
Best Local Similarity 25.7%; Pred. No. 0.00039;
Matches 69; Conservative 28; Mismatches 87; Indels 84; Gaps 9;

Qy 1 MKLVTKLGTHTFEIRVQPNDDTMAVKNIETEEQKDSYPMWQQLIFNGKVLKDESTLE 60
Db 126 MQIFVTKLTGKTTITLEVPSDFTIENVK---AKIQDEKIGIPDQQRILIFAGKLEDGRTLS 182

Qy 61 ENKVNEDG--FLVVMLSKGTSGTSTSSQHSNTPATROAPPLEAPQAPPPVAPITT 118
Db 183 DYNIQKESTLHLVLRGGM-----EQRSQH-----CKPEEGLEARGE- 221

Qy 119 SQPEGL-PAQAPNTHDNAASLLSGRNVDTTIINQLMWSGGSDKDKVQALRAAYNNPE 177
Db 222 ---LGLVGAQAPATEBQEAASSS-----TLVE----- 246

Qy 178 RAVEYLYSGIPVTAEIAVPIGGQANTTDRAPTEAGLSGIPNTAPDLFPQASNAGG 237
Db 247 -----VTLGEVPAESPPDPPOSQGASLPTTMYVPLMSQSVEDSSNQ 289

Qy 238 AGGGPLDFLRNNPQFA-----VREMWH 260
Db 290 EEEGPSTFPDLESEFQAALSRKVAELVH 317

RESULT 10
US-10-146-473-47
; Sequence 47, Application US/10146473
; Publication No. US2003010888A1
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Gout, Ivan
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Gure, Ali
; APPLICANT: Chen, Yao-Teeng
```

```

; APPLICANT: Old, Lloyd
; TITLE OF INVENTION: Breast Cancer Antigens
; FILE REFERENCE: L00461/70130(JRV)
; CURRENT APPLICATION NUMBER: US/10/146,473
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/291,150
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 47
; LENGTH: 624
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-146-473-47

Query Match
Best Local Similarity 21.0%; Score 150; DB 15; Length 624;
Matches 94; Conservative 61; Mismatches 157; Indels 136; Gaps 18;

Qy 1 MKLTVTKLGTHTFEIRVQPNNDITMAVKKNIEIQGKDSYPMWQQLLIFFNGKVLKDESTLE 60
Db 33 IKVTVPKEKE-EFAPVENSQVQFK--EAIKRFKSTQDLVLIFAGKILKQDQTLI 88

Qy 61 ENKVNEDGFLVMSKKG-----TSGST----- 83
Db 89 QHGIIH-DGLTVHLVIKSNRPQGSTQPSNAAAGTNTTSASTPRSNSTPISTNSNPFGLGS 147

Qy 84 -----GTSSQHSNTPATROAPPLEAPQ---QAPQPPVAPITTSQPEGL----- 124
Db 148 LGLAGLSSGLGSLSTFSELSQSQOQLMASPEMMIQIMENFVQSMLSNPDLMQLIMA 207

Qy 125 --PAQANTHDNAASLLSG-----RNVDTIINQLMEMGGGSD 161
Db 208 NPQOQQLIQRNPEISHLLNPNIDIMRTLEIARNPAMQEMMENQDLALSNSLESIPGG--- 264

Qy 162 KDKVQRALRAAYNPE-----RAVEYLYSGIPYTAETAVDIGQG---ANTTDRAPTGEAG 214
Db 265 ---YNALRRMYTDIQEPLNAAQEQFGNPFASVSGSSSGEGTQPSRTENRDP----- 315

Qy 215 LSGIPNT-APLDLDFPGA-----SNAGGAGGGGLDPLFN-----NPQFOAV 255
Db 316 ---LPNPWAPPPATQSSATSTTTTSGSGSGSSSNATGNTVAANYVASIPSTPGMOSL 372

Qy 256 REMVHTNPQIQPL-----VELSKQNPQILRLIENHDEFL--QLLNEPPEGGSGD 305
Db 373 LQQITENPQIQNLSAPYMRSMWQSLSQNPDLAAQMMLNSPLFTANPQLQEQMRPQLPA 432

Qy 306 FLDQPEDEMPHAISVTPPEQEAIGLE 333
Db 433 FLOQMOPNDTLSAMS-NPRAMQALMQIQ 459

RESULT 11
US-09-938-864-410
; Sequence 410, Application US/09938864
; Publication No. US20030072767A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Smithgall, Molly
; APPLICANT: Moulton, Gus
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Mossman, Sally
; APPLICANT: Evans, Lawrence
; APPLICANT: Spies, A. Gregory
; APPLICANT: Boydston, Jeremy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR WT1
; FILE REFERENCE: 210121.465C5
; CURRENT APPLICATION NUMBER: US/09/938,864
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 413

; SOFTWARE: Fast-SEQ for Windows Version 3.0
; SEQ ID NO 410
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-195-835-410

Query Match
Best Local Similarity 30.0%; Score 140; DB 12; Length 504;
Matches 48; Conservative 15; Mismatches 59; Indels 38; Gaps 4

Qy 1 MKLTVTKLGTHTFEIRVQPNNDITMAVKKNIEIQGKDSYPMWQQLLIFFNGKVLKDESTLE 60
Db 1 MQIFVTKLTGKTTITLEVFESDTIENVK---AKIQKEGIPPDQORLIFAGKQLEDRGRTLS 57

Qy 61 ENKVNEDGFLVVML-----SKGKTSSTGTSSTSSQHSNTPATROA 99
Db 58 DYNIQKESTLHLVRLRGAMGSDVRDLNALLPAVPSLGGGGCALPVSQAQWAPVLDFPA 117

Qy 100 PP-----LEAPQAPQPPVAPITTSQPEGLPAQAPN 130
Db 118 PPGASAYSLGGPAPPAPPPPPPP-----PHSFIKQEPS 152

RESULT 13
US-10-286-333-410
; Sequence 410, Application US/10286333
; Publication No. US20030215458A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Gaiger, Alexander
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Jaya, No. US20030215458A1alie
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR WT1
; TITLE OF INVENTION: SPECIFIC IMMUNOTHERAPY
; FILE REFERENCE: 210121.465C10
; CURRENT APPLICATION NUMBER: US/10/286,333
; CURRENT FILING DATE: 2002-10-30
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 410
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-286-333-410

Query Match      7.3%; Score 140; DB 12; Length 504;
Best Local Similarity 30.0%; Pred. No. 0.0052;
Matches 48; Conservative 15; Mismatches 59; Indels 38; Gaps 4;

QY 1 MKLTVKTLKGTHFEIRVQPNDTIMAVKNIEIQGKDSYPMGQQLIFNGKVLKDESTLE 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MQIFVKTLTGKTTITLEVEPSDTIENVK---AKIQDKEGIPDPQORLIFAGKQLEDGRTL 57
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 ENKVNEDGFLVVML-----SKGKTSGSTGTSSSQHSNTPATROA 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 DYNIOKESTLHLVLRGAMGSDVRDLNALLPAVPSLGGGGGCGALPVSGAAQWAPVLDF 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 100 PP-----LEAPQAPQPPVAPITTSQPEGLPAQAPN 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 PPGASAYGSLGGPAPPAPPPPPPP-----PHSFIKQEPS 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-10-125-635A-410
; Sequence 410, Application US/10125635A
; Publication No. US2003009635A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Smithgall, Molly D.
; APPLICANT: Carter, Darrick
; APPLICANT: Cheever, Martin A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Sutherland, R. Alec
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR WT1
; TITLE OF INVENTION: SPECIFIC IMMUNOTHERAPY
; FILE REFERENCE: 210121.465C7
; CURRENT APPLICATION NUMBER: US/10/125,635A
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 410
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-125-635A-410

Query Match      7.3%; Score 140; DB 15; Length 504;
Best Local Similarity 30.0%; Pred. No. 0.0052;
Matches 48; Conservative 15; Mismatches 59; Indels 38; Gaps 4;

QY 1 MKLTVKTLKGTHFEIRVQPNDTIMAVKNIEIQGKDSYPMGQQLIFNGKVLKDESTLE 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MQIFVKTLTGKTTITLEVEPSDTIENVK---AKIQDKEGIPDPQORLIFAGKQLEDGRTL 57
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 ENKVNEDGFLVVML-----SKGKTSGSTGTSSSQHSNTPATROA 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 DYNIOKESTLHLVLRGAMGSDVRDLNALLPAVPSLGGGGGCGALPVSGAAQWAPVLDF 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 100 PP-----LEAPQAPQPPVAPITTSQPEGLPAQAPN 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 PPGASAYGSLGGPAPPAPPPPPPP-----PHSFIKQEPS 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 15
US-10-002-603-410
; Sequence 410, Application US/10002603
; Publication No. US20030095971A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Smithgall, Molly
; APPLICANT: Moulton, Gus
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Mossman, Sally
; APPLICANT: Evans, Lawrence
; APPLICANT: Spies, A. Gregory
; APPLICANT: Boydston, Jeremy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR WT1
; TITLE OF INVENTION: SPECIFIC IMMUNOTHERAPY
; FILE REFERENCE: 210121.465C6
; CURRENT APPLICATION NUMBER: US/10/002,603
; CURRENT FILING DATE: 2001-10-30
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 410
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-002-603-410

Query Match      7.3%; Score 140; DB 15; Length 504;
Best Local Similarity 30.0%; Pred. No. 0.0052;
Matches 48; Conservative 15; Mismatches 59; Indels 38; Gaps 4;

QY 1 MKLTVKTLKGTHFEIRVQPNDTIMAVKNIEIQGKDSYPMGQQLIFNGKVLKDESTLE 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MQIFVKTLTGKTTITLEVEPSDTIENVK---AKIQDKEGIPDPQORLIFAGKQLEDGRTL 57
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 ENKVNEDGFLVVML-----SKGKTSGSTGTSSSQHSNTPATROA 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 DYNIOKESTLHLVLRGAMGSDVRDLNALLPAVPSLGGGGGCGALPVSGAAQWAPVLDF 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 100 PP-----LEAPQAPQPPVAPITTSQPEGLPAQAPN 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 PPGASAYGSLGGPAPPAPPPPPPP-----PHSFIKQEPS 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: December 17, 2003, 06:31:56
Job time : 127.158 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:17:26 ; Search time 129.014 Seconds  
(without alignments)  
2595.451 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 1910

Sequence: 1 MKLTVTLKGFHFIRVQPN.....CDRNEELAANYLLEHAGEED 368

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 5728757 seqs, 909918778 residues

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents\_AA\_Main.\*

1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097A\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US097B\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US099A\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US099B\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
28: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
29: /cgn2\_6/ptodata/1/paa/US103\_COMB.pep.\*  
30: /cgn2\_6/ptodata/1/paa/US104\_COMB.pep.\*  
31: /cgn2\_6/ptodata/1/paa/US106\_COMB.pep.\*  
32: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1910	100.0	368	21	US-09-708-427-51368 Sequence 51368, A

2	1910	100.0	368	23	US-09-805-550-4	Sequence 4, Appli
3	1910	100.0	368	32	US-60-109-728-4	Sequence 4, Appli
4	1792	93.8	246	21	US-09-708-427-51369	Sequence 51369, A
5	1540	80.6	295	21	US-09-708-427-51370	Sequence 51370, A
6	1524.5	79.8	369	30	US-10-437-963-161952	Sequence 161952, A
7	1288.5	67.5	252	28	US-10-219-999-49896	Sequence 49896, A
8	1288.5	67.5	289	30	US-10-425-114-59208	Sequence 59208, A
9	1185	62.0	363	30	US-10-424-599-170467	Sequence 170467, A
10	1169.5	61.2	371	19	US-09-513-996A-18372	Sequence 18372, A
11	1169.5	61.2	371	20	US-09-620-394B-6541	Sequence 6541, Ap
12	1168.5	61.2	363	30	US-10-424-599-219691	Sequence 219691, A
13	1153.5	60.4	379	22	US-09-791-537-55742	Sequence 55742, A
14	1115	58.4	368	19	US-09-513-996A-55531	Sequence 55531, A
15	1067.5	55.9	348	19	US-09-513-996A-18373	Sequence 18373, A
16	1067.5	55.9	348	20	US-09-620-394B-6542	Sequence 6542, Ap
17	1043.5	54.6	367	22	US-09-791-537-18661	Sequence 18661, A
18	1019	53.4	389	22	US-09-791-537-58054	Sequence 58054, A
19	1014	53.1	345	19	US-09-513-996A-55532	Sequence 55532, A
20	1013.5	53.1	400	28	US-10-219-999-35327	Sequence 35327, A
21	1013.5	53.1	400	30	US-10-424-599-256715	Sequence 256715, A
22	1013.5	53.1	400	32	US-60-324-109-20739	Sequence 20739, A
23	1013.5	53.1	433	30	US-10-425-114-45576	Sequence 45576, A
24	1006.5	52.7	403	30	US-10-424-599-200504	Sequence 200504, A
25	1004.5	52.6	392	28	US-10-219-999-43947	Sequence 43947, A
26	1004.5	52.6	425	30	US-10-425-114-54716	Sequence 54716, A
27	1003.5	52.5	392	30	US-10-424-599-256719	Sequence 256719, A
28	978	51.2	419	22	US-09-791-537-66481	Sequence 66481, A
29	978	51.2	419	24	US-09-935-625-8508	Sequence 8508, Ap
30	954	49.9	382	22	US-09-791-537-55721	Sequence 55721, A
31	931.5	48.8	378	19	US-09-513-996A-21981	Sequence 21981, A
32	931.5	48.8	378	19	US-09-513-996A-56723	Sequence 56723, A
33	912	47.7	402	28	US-10-219-999-62455	Sequence 62455, A
34	912	47.7	402	32	US-60-324-109-28561	Sequence 28561, A
35	912	47.7	402	32	US-60-324-109-28562	Sequence 28562, A
36	912	47.7	424	30	US-10-425-114-72275	Sequence 72275, A
37	908.5	47.6	405	28	US-10-219-999-39045	Sequence 39045, A
38	908.5	47.6	405	32	US-60-312-544-8573	Sequence 8573, Ap
39	908.5	47.6	405	32	US-10-219-999-50088	Sequence 50088, A
40	908.5	47.6	418	30	US-10-425-114-41386	Sequence 41386, A
41	908.5	47.6	418	30	US-10-425-114-58736	Sequence 58736, A
42	904	47.3	442	28	US-10-219-999-55396	Sequence 55396, A
43	904	47.3	442	30	US-10-425-114-64677	Sequence 64677, A
44	903	47.3	402	28	US-10-219-999-54143	Sequence 54143, A
45	903	47.3	424	30	US-10-425-114-59667	Sequence 59667, A

#### ALIGNMENTS

#### RESULT 1

US-09-708-427-51368  
; Sequence 51368, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: THEREBY  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708, 427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 51368  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1..368  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: 1..368  
; OTHER INFORMATION: Ceres Seq. ID 1926693  
US-09-708-427-51368

```
Query Match 100.0%; Score 1910; DB 21; Length 368;
Best Local Similarity 100.0%; Pred. No. 9.2e-149;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLTVKTLKGTHFEIRVQPNNDTMAVKKNIEEIQKDSYPWGQQLLIENGKVLKDESTLE 60
Db 1 MKLTVKTLKGTHFEIRVQPNNDTMAVKKNIEEIQKDSYPWGQQLLIENGKVLKDESTLE 60

Qy 61 ENKVNEDGLVVLMSKGTSGTSSQHSNTPATRQAPPLAPQAPQPPVAPITTSQ 120
Db 61 ENKVNEDGLVVLMSKGTSGTSSQHSNTPATRQAPPLAPQAPQPPVAPITTSQ 120

Qy 121 PEGLPQAQPNTHDAAASLLSGRNVDTTIINQLMEMGGSGWDKQVQRALRAAYNNPERAV 180
Db 121 PEGLPQAQPNTHDAAASLLSGRNVDTTIINQLMEMGGSGWDKQVQRALRAAYNNPERAV 180

Qy 181 EYLYSGIPVTAEIAPVIGGGANTTDRAPTGEAGLSGIPNTAPLDLFPQGASNAGGAGG 240
Db 181 EYLYSGIPVTAEIAPVIGGGANTTDRAPTGEAGLSGIPNTAPLDLFPQGASNAGGAGG 240

Qy 241 GPLDFLRNNPQFQAVREMHVTHNPQILQPMVLVELSKONPQILRLIEENHDEFLLQNEPFE 300
Db 241 GPLDFLRNNPQFQAVREMHVTHNPQILQPMVLVELSKONPQILRLIEENHDEFLLQNEPFE 300

Qy 301 GEGDFLDQPEDEMPHAI SVTPEEQEAI GRLESMDGDRARVIEAFACDRNEELAANYL 360
Db 301 GEGDFLDQPEDEMPHAI SVTPEEQEAI GRLESMDGDRARVIEAFACDRNEELAANYL 360

Qy 361 LEHAGEED 368
Db 361 LEHAGEED 368

RESULT 2
US-09-805-550-4
; Sequence 4, Application US/0980550
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Tagliani, Laura
; APPLICANT: Bowen, Benjamin
; APPLICANT: Wang, Xun
; TITLE OF INVENTION: Rad23 Genes and Uses Thereof
; FILE REFERENCE: 0964D
; CURRENT APPLICATION NUMBER: US/09/805,550
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 09/413,574
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: 60/109,728
; PRIOR FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-09-805-550-4

Query Match 100.0%; Score 1910; DB 23; Length 368;
Best Local Similarity 100.0%; Pred. No. 9.2e-149;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLTVKTLKGTHFEIRVQPNNDTMAVKKNIEEIQKDSYPWGQQLLIENGKVLKDESTLE 60
Db 1 MKLTVKTLKGTHFEIRVQPNNDTMAVKKNIEEIQKDSYPWGQQLLIENGKVLKDESTLE 60

Qy 61 ENKVNEDGLVVLMSKGTSGTSSQHSNTPATRQAPPLAPQAPQPPVAPITTSQ 120
Db 61 ENKVNEDGLVVLMSKGTSGTSSQHSNTPATRQAPPLAPQAPQPPVAPITTSQ 120

Qy 121 PEGLPQAQPNTHDAAASLLSGRNVDTTIINQLMEMGGSGWDKQVQRALRAAYNNPERAV 180
Db 121 PEGLPQAQPNTHDAAASLLSGRNVDTTIINQLMEMGGSGWDKQVQRALRAAYNNPERAV 180

Qy 181 EYLYSGIPVTAEIAPVIGGGANTTDRAPTGEAGLSGIPNTAPLDLFPQGASNAGGAGG 240
Db 181 EYLYSGIPVTAEIAPVIGGGANTTDRAPTGEAGLSGIPNTAPLDLFPQGASNAGGAGG 240

US-09-805-550-4

Query Match 100.0%; Score 1910; DB 32; Length 368;
Best Local Similarity 100.0%; Pred. No. 9.2e-149;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLTVKTLKGTHFEIRVQPNNDTMAVKKNIEEIQKDSYPWGQQLLIENGKVLKDESTLE 60
Db 1 MKLTVKTLKGTHFEIRVQPNNDTMAVKKNIEEIQKDSYPWGQQLLIENGKVLKDESTLE 60

Qy 61 ENKVNEDGLVVLMSKGTSGTSSQHSNTPATRQAPPLAPQAPQPPVAPITTSQ 120
Db 61 ENKVNEDGLVVLMSKGTSGTSSQHSNTPATRQAPPLAPQAPQPPVAPITTSQ 120

Qy 121 PEGLPQAQPNTHDAAASLLSGRNVDTTIINQLMEMGGSGWDKQVQRALRAAYNNPERAV 180
Db 121 PEGLPQAQPNTHDAAASLLSGRNVDTTIINQLMEMGGSGWDKQVQRALRAAYNNPERAV 180

Qy 181 EYLYSGIPVTAEIAPVIGGGANTTDRAPTGEAGLSGIPNTAPLDLFPQGASNAGGAGG 240
Db 181 EYLYSGIPVTAEIAPVIGGGANTTDRAPTGEAGLSGIPNTAPLDLFPQGASNAGGAGG 240

Qy 241 GPLDFLRNNPQFQAVREMHVTHNPQILQPMVLVELSKONPQILRLIEENHDEFLLQNEPFE 300
Db 241 GPLDFLRNNPQFQAVREMHVTHNPQILQPMVLVELSKONPQILRLIEENHDEFLLQNEPFE 300

Qy 301 GEGDFLDQPEDEMPHAI SVTPEEQEAI GRLESMDGDRARVIEAFACDRNEELAANYL 360
Db 301 GEGDFLDQPEDEMPHAI SVTPEEQEAI GRLESMDGDRARVIEAFACDRNEELAANYL 360

Qy 361 LEHAGEED 368
Db 361 LEHAGEED 368

RESULT 3
US-60-109-728-4
; Sequence 4, Application US/60109728
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Tagliani, Laura
; APPLICANT: Bowen, Benjamin
; APPLICANT: Wang, Xun
; TITLE OF INVENTION: Maize Rad23 Genes and Uses Thereof
; FILE REFERENCE: 0964P
; CURRENT APPLICATION NUMBER: US/60/109,728
; CURRENT FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-60-109-728-4

Query Match 100.0%; Score 1910; DB 32; Length 368;
Best Local Similarity 100.0%; Pred. No. 9.2e-149;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLTVKTLKGTHFEIRVQPNNDTMAVKKNIEEIQKDSYPWGQQLLIENGKVLKDESTLE 60
Db 1 MKLTVKTLKGTHFEIRVQPNNDTMAVKKNIEEIQKDSYPWGQQLLIENGKVLKDESTLE 60

Qy 61 ENKVNEDGLVVLMSKGTSGTSSQHSNTPATRQAPPLAPQAPQPPVAPITTSQ 120
Db 61 ENKVNEDGLVVLMSKGTSGTSSQHSNTPATRQAPPLAPQAPQPPVAPITTSQ 120

Qy 121 PEGLPQAQPNTHDAAASLLSGRNVDTTIINQLMEMGGSGWDKQVQRALRAAYNNPERAV 180
Db 121 PEGLPQAQPNTHDAAASLLSGRNVDTTIINQLMEMGGSGWDKQVQRALRAAYNNPERAV 180

Qy 181 EYLYSGIPVTAEIAPVIGGGANTTDRAPTGEAGLSGIPNTAPLDLFPQGASNAGGAGG 240
Db 181 EYLYSGIPVTAEIAPVIGGGANTTDRAPTGEAGLSGIPNTAPLDLFPQGASNAGGAGG 240

Qy 241 GPLDFLRNNPQFQAVREMHVTHNPQILQPMVLVELSKONPQILRLIEENHDEFLLQNEPFE 300
Db 241 GPLDFLRNNPQFQAVREMHVTHNPQILQPMVLVELSKONPQILRLIEENHDEFLLQNEPFE 300

Qy 301 GEGDFLDQPEDEMPHAI SVTPEEQEAI GRLESMDGDRARVIEAFACDRNEELAANYL 360
Db 301 GEGDFLDQPEDEMPHAI SVTPEEQEAI GRLESMDGDRARVIEAFACDRNEELAANYL 360

Qy 361 LEHAGEED 368
Db 361 LEHAGEED 368

RESULT 4
US-09-708-427-51369
; Sequence 51369, Application US/09708427
; GENERAL INFORMATION:
```



```

; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51369
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..345
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..345
; OTHER INFORMATION: Ceres Seq. ID 1926694
; US-09-708-427-51369

Query Match      93.8%; Score 1792; DB 21; Length 345;
Best Local Similarity 100.0%; Pred. No. 4.8e-139;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 MAVKNIIEIOGKSYPMGQQLIFNGKVLKDESTLEENKVNEDGFLVVLMSKGTSGST 83
Db 1 MAVKNIIEIOGKSYPMGQQLIFNGKVLKDESTLEENKVNEDGFLVVLMSKGTSGST 60

Qy 84 GTSSSOHNTATRAQAPPLAPQAPPPVAPITTSQEGPLPAQAPNTHDAAANLLSGR 143
Db 61 GTSSSOHNTATRAQAPPLAPQAPPPVAPITTSQEGPLPAQAPNTHDAAANLLSGR 120

Qy 144 NVDTIINQLMWGGSGWDKQVORALRAAYNNPERAVEYLISGIPVTAIEIAVPISGGGAN 203
Db 121 NVDTIINQLMWGGSGWDKQVORALRAAYNNPERAVEYLISGIPVTAIEIAVPISGGGAN 180

Qy 204 TTDRAFTGEAGLSGIPNTAPLDLPQGNASAGGAGGGLPFLRNNPQFQAVREMVHTNP 263
Db 181 TTDRAFTGEAGLSGIPNTAPLDLPQGNASAGGAGGGLPFLRNNPQFQAVREMVHTNP 240

Qy 264 QILQPMVLVSKONPQILRLIEENHDEFLLQNLNPFEGGEGDFLDQPEDEMPHAISVTP 323
Db 241 QILQPMVLVSKONPQILRLIEENHDEFLLQNLNPFEGGEGDFLDQPEDEMPHAISVTP 300

Qy 324 EEQEAGRLSGMGFDRARVIEAFACDRNEELAANYLLEHAGED 368
Db 301 EEQEAGRLSGMGFDRARVIEAFACDRNEELAANYLLEHAGED 345

RESULT 5
US-09-708-427-51370
; Sequence 51370, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51370
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..296
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..296
; OTHER INFORMATION: Ceres Seq. ID 1926695
; US-09-708-427-51370

Query Match      80.6%; Score 1540; DB 21; Length 296;
Best Local Similarity 100.0%; Pred. No. 2.6e-118;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 MLSKGTSGTSGTSSQHSNTPATRAQAPPLAPQAPPPVAPITTSQEGPLPAQAPNTH 132
Db 1 MLSKGTSGTSGTSSQHSNTPATRAQAPPLAPQAPPPVAPITTSQEGPLPAQAPNTH 60

Qy 133 DNAASNLISGRNVDTIINQLMWGGSGWDKQVORALRAAYNNPERAVEYLISGIPVTAIE 192
Db 61 DNAASNLISGRNVDTIINQLMWGGSGWDKQVORALRAAYNNPERAVEYLISGIPVTAIE 120

Qy 193 IAVPIGGQAGNTTDRAPTGEAGLSGIPNTAPLDLPQGNASAGGAGGGLPFLRNNPQF 252
Db 121 IAVPIGGQAGNTTDRAPTGEAGLSGIPNTAPLDLPQGNASAGGAGGGLPFLRNNPQF 180

Qy 253 QAVREMVHTNPQILQPMVLVSKONPQILRLIEENHDEFLLQNLNPFEGGEGDFLDQPEE 312
Db 181 QAVREMVHTNPQILQPMVLVSKONPQILRLIEENHDEFLLQNLNPFEGGEGDFLDQPEE 240

Qy 313 DEMPHAISVTPPEQEAIGRLSGMGFDRARVIEAFACDRNEELAANYLLEHAGED 368
Db 241 DEMPHAISVTPPEQEAIGRLSGMGFDRARVIEAFACDRNEELAANYLLEHAGED 296

RESULT 6
US-10-437-963-161952
; Sequence 161952, Application US/10437963
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yibua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 161952
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_6108C.1.pep
; US-10-437-963-161952

Query Match      79.8%; Score 1524.5; DB 30; Length 369;
Best Local Similarity 79.9%; Pred. No. 7e-117;
Matches 295; Conservative 31; Mismatches 42; Indels 1; Gaps 1;

Qy 1 MKLTVKTLKGTTFEIRVQPNDDTMAVKNIIEIOGKSYPMGQQLIFNGKVLKDESTLE 60
Db 1 MKLTVKTLKGTTFEIRVQPNDDTMAVKNIIEIOGKSYPMGQQLIFNGKVLKDESTLE 60

Qy 61 ENKVNEDGFLVVLMSKGTSGTSGTSSQHSNTPATRAQAPPLAPQAPPPVAPITTSQ 120
Db 61 ENKVNEDGFLVVLMSKGTSGTSGTSSQHSNTPATRAQAPPLAPQAPPPVAPITTSQ 120

Qy 121 PEGPLPAQAP-NTHDNAASNLISGRNVDTIINQLMWGGSGWDKQVORALRAAYNNPERA 179
Db 121 PERPPAEAPSNAAYQAAASNLISGRNVDTIINQLMWGGSGWDKQVORALRAAYNNPERA 180

Qy 180 VEYLISGIPVTAIEIAVPISGGQAGNTTDRAPTGEAGLSGIPNTAPLDLPQGNASAGGAG 239
Db 181 VEYLISGIPVTAIEIAVPISGGQAGNTTDRAPTGEAGLSGIPNTAPLDLPQGNASAGGAG 240

```

QY 240 GGPLDLFLRNPPQFAVRVMTNTPQILOPMLVELSKNPQILRLRIENHDEFLQLLNEPP 299  
Db 241 GGTLEFLRHNOQFALREVMVHTNPQILOPMLQELSKNPQILRLRIENHDEFLQLLNEPP 300  
QY 300 EGGEDGFLDQDEEMPHASIVTPEEQEAGRLSGMGFDRARVIEAFACDRNEELAANY 359  
Db 301 DGADGDFLDQDQDEMPHSINVTPEEQEAGRLSGMGFDRARVIEAFACDRNEOLAANY 360  
QY 360 LLEHAGEED 368  
Db 361 LLEHADED 369  
RESULT 7  
US-10-219-999-49896  
; Sequence 49896, Application US/10219999  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Edgerton, Michael D  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Stein, Joshua  
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-10(52726)C  
; CURRENT APPLICATION NUMBER: US/10/219,999  
; PRIOR FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: US 60/324,109  
; PRIOR FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/312,544  
; PRIOR FILING DATE: 2001-08-15  
; NUMBER OF SEQ ID NOS: 63520  
; SEQ ID NO 49896  
; LENGTH: 252  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-219-999-49896  
Query Match 67.5%; Score 1288.5; DB 28; Length 252;  
Best Local Similarity 99.2%; Pred. No. 1.3e-97;  
Matches 250; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Edgerton, Michael D  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: Stein, Joshua  
TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
FILE REFERENCE: 38-10(52726)C  
CURRENT APPLICATION NUMBER: US/10/219,999  
PRIOR FILING DATE: 2002-08-15  
PRIOR APPLICATION NUMBER: US 60/324,109  
PRIOR FILING DATE: 2001-09-21  
PRIOR APPLICATION NUMBER: US 60/312,544  
PRIOR FILING DATE: 2001-08-15  
NUMBER OF SEQ ID NOS: 63520  
SEQ ID NO 49896  
LENGTH: 252  
TYPE: PRT  
ORGANISM: Zea mays  
US-10-219-999-49896  
QY 1 MKLTVTKLKGTHFEIRVQPNNDTMAVKKNIEEIQGKDSYPWGQQLLIFNGKVLKDESTLE 60  
Db 1 MKLTVTKLKGTHFEIRVQPNNDTMAVKKNIEEIQGKDSYPWGQQLLIFNGKVLKDESTLE 60  
QY 61 ENKVNEDGFLVWMLSKGTSSTGTSQHSNTPATROAPPLEAPQOAPPPVAPITTSQ 120  
Db 61 ENKVNEDGFLVWMLSKGTSSTGTSQHSNTPATROAPPLEAPQOAPPPVAPITTSQ 120  
QY 121 PEGLPAPQAPNTHDAAASNLSSGRNVDTIINQLMEMGGGSKDKVQRALRAAYNNPERAV 180  
Db 121 PEGLPAPQAPNTHDAAASNLSSGRNVDTIINQLMEMGGGSKDKVQRALRAAYNNPERAV 180  
QY 181 EYLYSGIPVTAEIAVPIGGQGANTTDRAPTGEAGLSGIPNTAPLDLFP-QGASNAGGAG 239  
Db 181 EYLYSGIPVTAEIAVPIGGQGANTTDRAPTGEAGLSGIPNTAPLDLFP-QGASNAGGAG 239  
QY 240 GGPLDLFLRNPPQ 251  
Db 241 GGPLDLFLRNPPQ 252  
RESULT 8  
US-10-425-114-59208  
; Sequence 59208, Application US/10425114  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 59208  
; LENGTH: 289  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700264466\_FLI.pep  
US-10-425-114-59208  
Query Match 67.5%; Score 1288.5; DB 30; Length 289;  
Best Local Similarity 99.2%; Pred. No. 1.6e-97;  
Matches 250; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
QY 1 MKLTVTKLKGTHFEIRVQPNNDTMAVKKNIEEIQGKDSYPWGQQLLIFNGKVLKDESTLE 60  
Db 30 MKLTVTKLKGTHFEIRVQPNNDTMAVKKNIEEIQGKDSYPWGQQLLIFNGKVLKDESTLE 89  
QY 61 ENKVNEDGFLVWMLSKGTSSTGTSQHSNTPATROAPPLEAPQOAPPPVAPITTSQ 120  
Db 90 ENKVNEDGFLVWMLSKGTSSTGTSQHSNTPATROAPPLEAPQOAPPPVAPITTSQ 149  
QY 121 PEGLPAPQAPNTHDAAASNLSSGRNVDTIINQLMEMGGGSKDKVQRALRAAYNNPERAV 180  
Db 150 PEGLPAPQAPNTHDAAASNLSSGRNVDTIINQLMEMGGGSKDKVQRALRAAYNNPERAV 209  
QY 181 EYLYSGIPVTAEIAVPIGGQGANTTDRAPTGEAGLSGIPNTAPLDLFP-QGASNAGGAG 239  
Db 210 EYLYSGIPVTAEIAVPIGGQGANTTDRAPTGEAGLSGIPNTAPLDLFP-QGASNAGGAG 269  
QY 240 GGPLDLFLRNPPQ 251  
Db 270 GGPLDLFLRNPPQ 281  
RESULT 9  
US-10-424-599-170467  
; Sequence 170467, Application US/10424599  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 170467  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(363)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_124948C.1.pap  
US-10-424-599-170467  
Query Match 62.0%; Score 1185; DB 30; Length 363;  
Best Local Similarity 64.0%; Pred. No. 8.2e-89;  
Matches 236; Conservative 51; Mismatches 70; Indels 12; Gaps 5;  
QY 1 MKLTVTKLKGTHFEIRVQPNNDTMAVKKNIEEIQGKDSYPWGQQLLIFNGKVLKDESTLE 60  
Db 1 MKLTVTKLKGTHFEIRVQPNNDTMAVKKNIEEIQGKDSYPWGQQLLIFNGKVLKDESTLE 60



APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 219691  
LENGTH: 363  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_40409C.1.pep  
US-10-424-599-219691

Query Match 61.2%; Score 1168.5; DB 30; Length 363;  
Best Local Similarity 63.1%; Pred. No. 1.9e-87;  
Matches 236; Conservative 51; Mismatches 70; Indels 17; Gaps 6;

QY 1 MKLTVKTLKGSHFEIRVQPNQDITMAVKNIIEIQGKDSYPMQOQLLIHNGKVLKDESTLE 60  
DB 1 MKLTVKTLKGSHFEIRVQPNQDITMAVKNIIEIQGKDSYPMQOQLLIHNGKVLKDESTLV 60

QY 61 ENKVNEDGFLVWMLSGKTSSTGTSSTSSOHSNTPAT-----RQAPLEAPQAPPPVAP 115  
DB 61 ENKVNEDGFLVWMLSGKTSSTGTSSTSSOHSNTPAT-----RQAPLEAPQAPPPVAP 120

QY 116 ITTSOPEGLPAQPNTHNAASNLISGRNVDTIINQMEMGGSGWDDKQVQALRAAYNN 175  
DB 121 SSTADP--TTNVSADTYGLAASNLVAGSNLEQTIQIIMDMGGNWDRTVSRALRAAYNN 178

QY 176 PERAVEYLXGIPVTAEIAVPIG--GOGANTTDRAPTEAGLSGIPNTAPLDLFPQASNAG 235  
DB 179 PERAIDLYLSGIPAEAAVAVPQTAGISSGAVPVG-----PNSPLNMFQO-ETISS 230

QY 236 GGAGGGLFLRNPFQAVREVMVHTNPQILQPMVLVELSKONPQILRLIEENHDFLQLL 295  
DB 231 TGAGLSGLFLRNPFQALRSVQSNPQLQVQLQELGKQNPGLRLIQEHGHEFLQI 290

QY 296 NEPFEGEGDFLDQPEDEMPHAIISVTPEQEAIGRLSMGDFDRARVIEAFACDRNEEL 355  
DB 291 NEPFEGEGDFLDQPEDEMPHAIISVTPEQEAIGRLSMGDFDRARVIEAFACDRNEEL 349

QY 356 AANYLLEHAGE-ED 368  
DB 350 AANYLLENAGDFED 363

RESULT 13  
US-09-791-537-55742  
Sequence 55742, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Bionomix, Inc.  
APPLICANT: Debe, Derek  
TITLE OF INVENTION: Danzer, Joseph  
FILE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 55742  
LENGTH: 379  
TYPE: PRT  
ORGANISM: Daucus carota  
US-09-791-537-55742

Query Match 60.4%; Score 1153.5; DB 22; Length 379;  
Best Local Similarity 63.0%; Pred. No. 3.5e-86;  
Matches 240; Conservative 45; Mismatches 81; Indels 15; Gaps 9;

QY 1 MKLTVKTLKGTHFEIRVQPNQDITMAVKNIIEIQGKDSYPMQOQLLIHNGKVLKDESTLE 60

DB 1 MKLTVKTLKGSHFEIRVQPNQDITMAVKNIIEIQGKDSYPMQOQLLIHNGKVLKDESTLA 60  
QY 61 ENKVNEDGFLVWMLSGKTSSTGTSSTSSOHSNTPATQAP---PLEAPQAPQPPVAPIT 117  
DB 61 ESKISEDGFLVWMLSGKTSSTGTPAAQSSAPAPTPAPAVAPAPAPAPASAVIPNT 120

QY 118 TSQPEG--LPAQAP-NTHNAASNLISGRNVDTIINQMEMGGSGWDDKQVQALRAAYNN 174  
DB 121 T-VPEAPLSFAFAPSDTYGEAASNVAGSNLEQTIQIIMDMGGNWDRTVSRALRAAYNN 179

QY 175 NPERAVEYLXGIPVTAEIAVPIG--GOGANTTDRAPTEAGLSGIPNTAPLDLFP 228  
DB 180 NPERAVEYLXGIPVTAEIAVPIG--GOGANTTDRAPTEAGLSGIPNTAPLDLFP 239

QY 229 QGASNAGGAGGGLFLRNPFQAVREVMVHTNPQILQPMVLVELSKONPQILRLIEENH 288  
DB 240 QETLSGVYTGAGLSGLFLRNPFQAVREVMVHTNPQILQPMVLVELSKONPQILRLIEENH 299

QY 289 DEFLQLLNEPPEGEGDFLDQPEDEMPHAIISVTPEQEAIGRLSMGDFDRARVIEAFLA 348  
DB 300 EEFQLQINPEPVEASEGDMFQDQEQD-VPEQITVTAADQEAIERLEAMGDFDRGLVIEAFLA 358

QY 349 CDRNEELAANYLLEHAGE-ED 368  
DB 359 CDRNEELAVNYLLENAGDFED 379

RESULT 14  
US-09-513-996A-55531  
Sequence 55531, Application US/09513996A  
GENERAL INFORMATION:  
APPLICANT: N. ALEXANDROV et al.  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDE:  
FILE OF INVENTION: ENCODED THEREBY  
FILE REFERENCE: 2750-709P  
CURRENT APPLICATION NUMBER: US/09/513,996A  
CURRENT FILING DATE: 2000-02-25  
NUMBER OF SEQ ID NOS: 81028  
SEQ ID NO 55531  
LENGTH: 368  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: 1..368  
OTHER INFORMATION: any n or Xaa = unknown  
FEATURE:  
OTHER INFORMATION: Location 1..368 / Ceres Seq. ID 2113785  
US-09-513-996A-55531

Query Match 58.4%; Score 1115; DB 19; Length 368;  
Best Local Similarity 62.8%; Pred. No. 5.1e-83;  
Matches 235; Conservative 45; Mismatches 82; Indels 12; Gaps 8;

QY 1 MKLTVKTLKGTHFEIRVQPNQDITMAVKNIIEIQGKDSYPMQOQLLIHNGKVLKDESTLE 60  
DB 1 MKLTVKTLKGSHFEIRVQPNQDITMAVKNIIEIQGKDSYPMQOQLLIHNGKVLKDESTLV 60

QY 61 ENKVNEDGFLVWMLSGKTSSTGTSSTSSOHSNTPATQAPLEAPQAPQPPVAPITTS 119  
DB 61 ENKVNEDGFLVWMLSGKTSSTGTSSTSSOHSNTPATQAPLEAPQAPQPPVAPITTS 119

QY 120 QP-EGPAPQAPNTHNAASNLISGRNVDTIINQMEMGGSGWDDKQVQALRAAYNNPER 178  
DB 120 TPVQEQPTAQSDTYGQAASLVSGSSIEQMVQOIMEMGGSGWDDKQVQALRAAYNNPER 179

QY 179 AVEYLXGIPVTAEI-AVPIGOGANTTDRAPTEAGLSGIPNTAPLDLFPQASNAGG 237  
DB 180 ADVLYLSGIPETVTPATNLVSGSGRELTAPPP-----SGGNSSPLDLFPQASVDAAG 235

QY 238 AGGGGLDFLRNPFQAVREVMVHTNPQILQPMVLVELSKONPQILRLIEENHDFLQLINE 297

Dd		236	GDLGTLFLRGNQDQQQLRSWVNSNPQTLPMLQELKQNPKLRLIQENQAEBFQLINE	295
Qy		298	PFEGBEG--DFLDQBEEDEMPHATSVTPPEQBATGRLESMSGFDRARVFIEAFACDNREBL	355
	:		: : :	
Dd		296	PYSGDGVDITFDQPQ-EMPHSNVTPEEQESTERLEANGFPDAIVIEAFLSCDRNEB	354
	:		: : :	
Qy		356	AANYLLRHAGE-ED	368
	:		: : :	
Dd		355	AANYLLHSADPED	368
	:		: : :	

**RESULT 15**

```

US-09-513-996A-18373
; Sequence 18373, Application US/09513996A
; GENERAL INFORMATION:
; TITLE: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 18373
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..348
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..348 / Ceres Seq. ID 1396370
US-09-513-996A-18373

```

Query Match	55.9%;	Score 1067.5;	DB 19;	Length 348;
Best Local Similarity	62.9%;	Pred. No. 4e-79;		
Matches 227;	Conservative 41;	Mismatches 64;	Indels 29;	Gaps 9;
QY	24	MAVKNIETEEIQKDSYPMWGQQLIFNGKVLKDBSTLEENKVNEDGFLVLMLSKGTSGST	83	
Db	1	MAVKNIETDSQKDNYPGCGQLLIHNGKVLKDBSTLVENKVTGEGFLVLMLSKSGSGSA	60	
QY	84	GTSSSQHNS-----TPATROAPPLEAPQQAQPPVAPITTTSQPEGLPAQAPNTHDNA	135	
Db	61	GQASVQTSSVSQPVSAATSTTKPAAP--STTQSPVPASPIPAQEQ--FAAQTDITYGQA	115	
QY	136	ASNLLSGRVDTIINQIMEMGGSSWDKDKVQRALRAAYNNPERAVEVLYSGIPVTABIAV	195	
Db	116	ASTLVSGSSLEQVMQQIMEMGGSSWDKXETVTRALRAAYNNPERAVOVLVYSGIPQTAEVAV	175	
QY	196	P-----IGGGCANTTDRAPTGEAGLSGIPNTAPLDLFPQGSNAGGAGGPGDLDLFRNNP	250	
Db	176	VPVPEAQIAGSGA-----APVAPA--SGGPNSSPLDLFPQETVAAGSGDLGTLEFLRND	228	
QY	251	QFQAVREWVTNQLLOPMLVELSKQNPOILRLIEENHDEFLOLLNPPFG--GEGDFLD	308	
Db	229	QFQQLRTMWSNFIQLQPMLOELGKQNPOLLRLIQENQAEFLQVNEPFYSGSGEGDMFD	288	
QY	309	QPQEDEMPHAI SVTPPEQEAIGRLMSGDFRARIETAEFLACDRNEELAAVNYLLEHAGE-E	367	
Db	289	QPQEQ-EMPHAINVTPAEQEAIQRLQLEAMGFDRALVIEFLACDRNEELAAVNYLLENSGDFF	347	
QY	368	D 368		
Db	348	D 348		

Search completed: December 17, 2003, 06:26:01  
Job time : 130.014 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:16:40 ; Search time 13.3299 Seconds  
(without alignments)  
1168.082 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 1910

Sequence: 1 MKLTVKTLKGTHTFEIRVQPN.....CDRNEELAANYLLEHAGEED 368

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1910	100.0	368	3	US-09-413-574-4
2	900.5	47.1	405	3	US-09-413-574-2
3	182	9.5	79	3	US-09-100-802-4
4	173.5	9.1	126	4	US-08-747-562-26
5	164	8.6	81	3	US-09-100-802-5
6	154	8.1	915	2	US-08-480-917-2
7	154	8.1	915	3	US-09-138-736-2
8	154	8.1	915	4	US-08-988-242-2
9	133.5	7.0	1274	4	US-09-095-443-2
10	126.5	6.6	156	2	US-09-070-060-7
11	126.5	6.6	156	3	US-09-051-969A-3
12	126.5	6.6	156	3	US-09-051-969A-4
13	126.5	6.6	156	3	US-09-357-746-7
14	126.5	6.6	160	4	US-09-370-838-205
15	126.5	6.6	229	2	US-08-726-306A-23
16	126.5	6.6	229	3	US-08-840-146-20
17	126.5	6.6	229	3	US-09-360-220-20
18	126	6.6	2616	6	5206163-3
19	125.5	6.6	198	4	US-08-988-242-19
20	125.5	6.6	352	3	US-08-854-764-2
21	125.5	6.6	352	5	PCT-US95-09377-2
22	123.5	6.5	76	3	US-08-817-787-8
23	122.5	6.4	76	2	US-09-070-060-8
24	122.5	6.4	76	3	US-09-357-746-8
25	122.5	6.4	76	3	US-08-817-787-1
26	122.5	6.4	76	3	US-09-100-802-1
27	122.5	6.4	76	4	US-09-331-930A-26

28 122.5 6.4 103 2 US-08-771-201-9 Sequence 9, Appli  
29 122.5 6.4 114 2 US-08-771-201-10 Sequence 10, Appli  
30 122.5 6.4 147 2 US-08-771-201-11 Sequence 11, Appli  
31 121 6.3 1121 1 US-07-789-915A-2 Sequence 2, Appli  
32 121 6.3 1121 1 US-08-005-002C-2 Sequence 2, Appli  
33 121 6.3 1121 1 US-08-487-203A-2 Sequence 2, Appli  
34 119.5 6.3 323 2 US-08-747-788-2 Sequence 2, Appli  
35 119.5 6.3 323 4 US-09-300-681B-2 Sequence 2, Appli  
36 117.5 6.2 457 4 US-09-516-277-2 Sequence 2, Appli  
37 117.5 6.2 457 4 US-09-516-277-4 Sequence 2, Appli  
38 117.5 6.2 533 1 US-08-462-092-2 Sequence 2, Appli  
39 117.5 6.2 533 3 US-08-746-822-2 Sequence 2, Appli  
40 117.5 6.2 533 3 US-09-094-350-2 Sequence 2, Appli  
41 117.5 6.2 533 6 5510474-2 Patent No. 5510474  
42 116 6.1 447 4 US-09-480-921B-6 Sequence 6, Appli  
43 115 6.0 158 4 US-09-091-725-25 Sequence 25, Appli  
44 114.5 6.0 76 1 US-08-232-815-2 Sequence 2, Appli  
45 114.5 6.0 76 1 US-08-350-906-2 Sequence 2, Appli

#### ALIGNMENTS

RESULT 1  
US-09-413-574-4  
; Sequence 4, Application US/09413574  
; Patent No. 6235972  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; APPLICANT: Tegliani, Laura  
; TITLE OF INVENTION: Maize Rad23 Genes and Uses Thereof  
; FILE REFERENCE: 0964  
; CURRENT APPLICATION NUMBER: US/09/413,574  
; CURRENT FILING DATE: 1999-10-06  
; EARLIER APPLICATION NUMBER: 60/109,728  
; EARLIER FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-413-574-4

Query Match 100.0%; Score 1910; DB 3; Length 368;  
Best Local Similarity 100.0%; Pred No. 8.3e-164;  
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLTVKTLKGTHTFEIRVQPNNDTMAVKKNIEEIQGKDSYPMGQQLLIFNGKVLKDESTLE 60  
Db 1 MKLTVKTLKGTHTFEIRVQPNNDTMAVKKNIEEIQGKDSYPMGQQLLIFNGKVLKDESTLE 60

Qy 61 ENKVNEDGFLVWLSKGTSGTGTSSQHSNTPATRAQAPPLPAPQAPPPVAPITTSQ 120  
Db 61 ENKVNEDGFLVWLSKGTSGTGTSSQHSNTPATRAQAPPLPAPQAPPPVAPITTSQ 120

Qy 121 PEGLPQAQPNTHNAASNLSSGRNVTIINQLMEMGGSGMDKQVORALRAAYNNPERAV 180  
Db 121 PEGLPQAQPNTHNAASNLSSGRNVTIINQLMEMGGSGMDKQVORALRAAYNNPERAV 180

Qy 181 EYLISGIPVTAETAVPIGGQAGNTTDRAPTGEAGLSGIPNTAPLDLFPQGSNAGGAGG 240  
Db 181 EYLISGIPVTAETAVPIGGQAGNTTDRAPTGEAGLSGIPNTAPLDLFPQGSNAGGAGG 240

Qy 241 GPLDLFRNNPQOAVREMVHTNPQIQLPMLVELSKQNPQILRIENHDFLQLLNEPPE 300  
Db 241 GPLDLFRNNPQOAVREMVHTNPQIQLPMLVELSKQNPQILRIENHDFLQLLNEPPE 300

Qy 301 GGEGLDQDPEEDMEPHATSVTPEEQEATCELSMGDFDRARVTEAFACDRNEELANYL 360  
Db 301 GGEGLDQDPEEDMEPHATSVTPEEQEATCELSMGDFDRARVTEAFACDRNEELANYL 360

Qy 361 LEHAGSED 368

Wed Dec 17 06:49:42 2003

```
Db 361 LEHAGEED 368
|||||
RESULT 2
US-09-413-574-2
; Sequence 2, Application US/09413574
; Patent No. 6235972
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Taghian, Laura
; FILE OF INVENTION: Maize Rad23 Genes and Uses Thereof
; FILE REFERENCE: 0964
; CURRENT APPLICATION NUMBER: US/09/413,574
; CURRENT FILING DATE: 1999-10-06
; EARLIER APPLICATION NUMBER: 60/109,728
; EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Zea mays
US-09-413-574-2
Query Match 47.1%; Score 900.5; DB 3; Length 405;
Best Local Similarity 49.0%; Pred. No. 7e-73;
Matches 197; Conservative 53; Mismatches 117; Indels 35; Gaps 7;
Qy 1 M K L V K T L K G T H F E I R V Q P N D T I M A V K N I E I Q G K D S Y P W G Q Q L L I F N G K V L K D E S T L E 60
Db 1 M K L N V K T L K G T N F E I E A S P D A S V A D V K R I E T T Q G S T Y R A D Q O M L I Y Q G K I L K D E T T L E 60
Qy 61 E N K V N E D G F L V M L S K G K T S G T S S S O H S N T P A T R A P ----- P L E A P Q A P - Q P P V A 114
Db 61 S N G V A E N S F L V M L S K A K A S - S G A S T A T T A K A P A T L A Q P A P V A P A A S V A R T P T Q A P V A 119
Qy 115 P I T T S O P E G L P A Q A P ----- N T H D N A A S N L S G R N V D T I I N Q L M E M G G S W D K D K 164
Db 120 T A E T A P S V Q P Q A P A A T V A A T D D A D V Y S Q A S N L V F G N N S Q T I Q I L D M G G T W E R D T 179
Qy 165 V Q R A L R A A Y N N P E R A V E Y I L Y S I P V T A E ----- I A V P I G G G A N T T D R A P T G E A ----- 213
Db 180 V V R A L R A A Y N N P E R A I D Y L Y S I P E N V E A P V A R A P A A G Q O T N Q A S P A Q P A V A L P V Q P 239
Qy 214 - G L S G I P N T A P L D F P Q G A S N A G - ----- G A G G P L D F L R N N P O F Q A V R E M V H T N P Q I L 266
Db 240 S P A S A G F N A N L P F P Q G V P S G S G N F G V P G A G S G A L D A R L Q P F Q A L L Q V A N P Q I L 299
Qy 267 Q P M L V E L S K O N P O I L R L I E N H D E F L Q L N E P P E G G E G D F L D Q P E D E M P H A I S V T P P E Q 326
Db 300 Q P M Q L E G K O N P O I L R L I Q E N Q A E F L V N E S P E G G P G N I L Q L A A V P Q T I L V T P P E R 359
Qy 327 E A I G R L E S M G F D R A R V I E A F L A C R N E E L A A N Y L L E H A G E E D 368
Db 360 E A I Q R L E G M G F N E L V I E V F F A C K D E E L T A N Y L L D H G H E F D 401
RESULT 3
US-09-100-802-4
; Sequence 4, Application US/09100802A
; Patent No. 6294363
; GENERAL INFORMATION:
; APPLICANT: Madura, Kiran
; TITLE OF INVENTION: Methods and Compositions for the Rapid
; TITLE OF INVENTION: Purification of Proteasomes and Methods of Use of Components
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: UMDN97-11
; CURRENT APPLICATION NUMBER: US/09/100,802A
; CURRENT FILING DATE: 1998-06-19
; EARLIER APPLICATION NUMBER: 60/050,171
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 17
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Sequence
US-09-100-802-4
Query Match 9.5%; Score 182; DB 3; Length 79;
Best Local Similarity 43.6%; Pred. No. 2.9e-09;
Matches 34; Conservative 23; Mismatches 21; Indels 0; Gaps 0;
Qy 1 M K L V K T L K G T H F E I R V Q P N D T I M A V K N I E I Q G K D S Y P W G Q Q L L I F N G K V L K D E S T L E 60
Db 1 M Q V T L K T L Q O Q T F K I D I D P E E T V K A L K E I S E K G D A F P V A G Q K L I Y A G K I L M D D T A L K 60
Qy 61 E N K V N E D G F L V M L S K G K 78
Db 61 E Y K I D E K N F V V M V T K P K 78
RESULT 4
US-08-747-562-26
; Sequence 26, Application US/08747562
; Patent No. 6579897
; GENERAL INFORMATION:
; APPLICANT: WALLACH, David
; APPLICANT: BOLDIN, Mark
; APPLICANT: METT, Igor
; APPLICANT: VARFOLOMEV, Eugene
; TITLE OF INVENTION: MODULATOR OF TNF/NGF SUPERFAMILY RECEPTORS
; TITLE OF INVENTION: AND SOLUBLE OLIGOMERIC TNF/NGF SUPERFAMILY RECEPTORS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,562
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05854
; FILING DATE: 11-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 109,632
; FILING DATE: 11-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 111,125
; FILING DATE: 02-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: WALLACH=15A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-747-562-26
```

```

Query Match          9.1%; Score 173.5; DB 4; Length 126;
Best Local Similarity 33.3%; Pred. No. 3.3e-08;
Matches 38; Conservative 25; Mismatches 32; Indels 19; Gaps 3;

Qy 13 FEIRVQPDNTMAVKNTIEIQKDSYPWGQQLIFNGKVLKDESTLEENKVNEDGFLAV 72
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 2 FKIRMEPDTVKLEKEIAEKGRDAPFVAGQKLIYAGKILSDDDPIRDYRIDEKNFVV 61
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 73 MLSKKTSGTGTSSQSHNTPATRQAPPLPAQOAPQ-----PPVAPITTSQP 121
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 62 MVT--KTRAGQOTS-----APPEASPTAAPESSTSFPPAPTSGMSHP 101
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 5
US-09-100-802-5
; Sequence 5, Application US/09100802A
; Patent No. 6294363
; GENERAL INFORMATION:
; APPLICANT: Madura, Kiran
; TITLE OF INVENTION: Methods and Compositions for the Rapid
;   TITLE OF INVENTION: Purification of Proteasomes and Methods of Use of Components
;   TITLE OF INVENTION: Thereof
; FILE REFERENCE: UMDNJ97-11
; CURRENT APPLICATION NUMBER: US/09/100,802A
; CURRENT FILING DATE: 1998-06-19
; EARLIER APPLICATION NUMBER: 60/050,171
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Sequence
US-09-100-802-5

Query Match          8.6%; Score 164; DB 3; Length 81;
Best Local Similarity 38.2%; Pred. No. 1.2e-07;
Matches 29; Conservative 26; Mismatches 21; Indels 0; Gaps 0;

Qy 3 LTVTKLKTHTHEIRVQPDNTMAVKNTIEIQKDSYPWGQQLIFNGKVLKDESTLEEN 62
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 5 ITLKTLOQOTFKIRMEPDTVKLEKEIAEKGRDAPFVAGQKLIYAGKILSDDDPIRDY 64
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 63 KVNEDGFLVWMLSKGK 78
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 65 RIDENKVVVWVTKK 80
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 6
US-08-480-917-2
; Sequence 2, Application US/08480917
; Patent No. 5820864
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucaia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: NEW TRYPAOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
;   TITLE OF INVENTION: CHAGAS DISEASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/138,736

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,917
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36400
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "protein"
US-08-480-917-2

Query Match          8.1%; Score 154; DB 2; Length 915;
Best Local Similarity 22.9%; Pred. No. 3.6e-05;
Matches 56; Conservative 43; Mismatches 126; Indels 20; Gaps 6;

Qy 57 STLEENKVNEDGFLVWMLSKGKTSVGTSGTSSQSHNTPATRQAPPLPAQOAPQPPVAPI 116
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 356 SRREEQPGQKTSVATAKPGCVSSGTDAASSHNTTSSAAASPPSPVSAKAAAP- 414
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 117 TTSQPEGLPAQANTHDNAASNLGRNVDTIINQLMEMGGGWDKQVQALRAAYNNP 176
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 415 ----PAAARSAPHVGSKIIANLVNQLGINVTQSVVSTGAPATR--STAVTSTTAP 467
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 177 ERAVEYLISGIPYTAETAVPIGGGANTTDRA---PTGEAGLSGIPNTAPLDLPQGSN 233
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 468 QRTSPYGHNGRPVTAGLVAANSAGSAASPTAAAKPTGEEKSAACSTSSVAINATRPAL 527
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 234 AGGAGAGGPDLDLRNNPQFQAVREVMHTNPQILQPMVELSKQNPQILRL----IENHD 289
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 528 HNASLPQAPTDGVLAAAVYQSEGE-VHQSLESLVITNSTR-----VLKLLPDTIRRDHE 582
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 290 EFLQL 294
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 583 QLLNL 587
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 7
US-09-138-736-2
; Sequence 2, Application US/09138736
; Patent No. 6270767
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucaia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: NEW TRYPAOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
;   TITLE OF INVENTION: CHAGAS DISEASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/138,736

```



; FILING DATE: US/08/988,242  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 30,024  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berridge, William P.  
; REGISTRATION NUMBER: 30,024  
; REFERENCE/DOCKET NUMBER: WPB 36400A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6400  
; TELEFAX: 703-836-2787  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: /desc = "protein"  
; DESCRIPTION: /desc = "protein"  
; US-08-988-242-2

Query Match 8.1%; Score 154; DB 4; Length 915;  
Best Local Similarity 22.9%; Pred. No. 3.6e-05;  
Matches 56; Conservative 43; Mismatches 126; Indels 20; Gaps 6;

QY 57 STLENNKVEDGLVLMVLSKGTSGTSSQHSNTPATROAPLEAPQAPPPVAPI 116  
Db 356 SRREQQPGQKTSVVATAKPGCVSSGTDASSSHNTTSAASASPSPVSAKAAAP- 414  
QY 117 TTSQEGPLPAQANTHDNAASNLISGRNVDTIINQLEMGGGSDKDKVQALRAAYNP 176  
Db 415 ----PAAARSAPHVGSKIIVNLGINVTQSVSTGAPATTR---STAVTSTTAP 467  
QY 177 ERAVELYSGIPVTAETIAVPIGGGANTTDRA---PTGEAGLSGIPNTAPLDLFPQASN 233  
Db 468 QRTSPYGHNGRPVTAGIVAANSASASPPTAAAKPTGEKASACETSSVAINATRPAL 527  
QY 234 AGGAGGGPLDPLRNPPQFQAVREMVHTNPQILQPLMLVELSKNQPIILRL-----IEENHD 289  
Db 528 HNASLPQAPTDGLVLAAYVQSEGE-VHQSLESLRVITNTSR-----VLKLLPDTIRRDHE 582  
QY 290 EFLQL 294  
Db 583 QLLNL 587

RESULT 9  
US-09-095-443-2  
; Sequence 2, Application US/09095443  
; Patent No. 6342593  
; GENERAL INFORMATION:  
; APPLICANT: Plowman, Gregory  
; APPLICANT: Peles, Eior  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT  
; TITLE OF INVENTION: OF ALP RELATED DISORDERS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSeq for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/095,443  
; FILING DATE: Herewith  
; CLASSIFICATION:

; FILING DATE: US 08/480,917  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 30,024  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berridge, William P.  
; REGISTRATION NUMBER: 30,024  
; REFERENCE/DOCKET NUMBER: WPB 36400  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6400  
; TELEFAX: 703-836-2787  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "protein"  
; US-09-138-736-2

Query Match 8.1%; Score 154; DB 3; Length 915;  
Best Local Similarity 22.9%; Pred. No. 3.6e-05;  
Matches 56; Conservative 43; Mismatches 126; Indels 20; Gaps 6;

QY 57 STLENNKVEDGLVLMVLSKGTSGTSSQHSNTPATROAPLEAPQAPPPVAPI 116  
Db 356 SRREQQPGQKTSVVATAKPGCVSSGTDASSSHNTTSAASASPSPVSAKAAAP- 414  
QY 117 TTSQEGPLPAQANTHDNAASNLISGRNVDTIINQLEMGGGSDKDKVQALRAAYNP 176  
Db 415 ----PAAARSAPHVGSKIIVNLGINVTQSVSTGAPATTR---STAVTSTTAP 467  
QY 177 ERAVELYSGIPVTAETIAVPIGGGANTTDRA---PTGEAGLSGIPNTAPLDLFPQASN 233  
Db 468 QRTSPYGHNGRPVTAGIVAANSASASPPTAAAKPTGEKASACETSSVAINATRPAL 527  
QY 234 AGGAGGGPLDPLRNPPQFQAVREMVHTNPQILQPLMLVELSKNQPIILRL-----IEENHD 289  
Db 528 HNASLPQAPTDGLVLAAYVQSEGE-VHQSLESLRVITNTSR-----VLKLLPDTIRRDHE 582  
QY 290 EFLQL 294  
Db 583 QLLNL 587

RESULT 8  
US-08-988-242-2  
; Sequence 2, Application US/08988242  
; Patent No. 6403103  
; GENERAL INFORMATION:  
; APPLICANT: PARANHOS-BACCALA, GLAUCIA  
; APPLICANT: LESENECHAL, MYLENE  
; APPLICANT: JOLIVET, MICHEL  
; APPLICANT: MANBRAND, BERNARD  
; TITLE OF INVENTION: NEW TRYPAANOSOMA CRUZI ANTIGEN, GENE  
; TITLE OF INVENTION: ENCODING THEREFOR, AND METHODS OF DETECTING AND TREATING  
; TITLE OF INVENTION: CHAGAS DISEASE  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OLIFF & BERRIDGE, PLC  
; STREET: P.O. BOX 19928  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/049,477
; FILING DATE: June 12, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 235/055
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1274 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-095-443-2

Query Match 7.0%; Score 133.5; DB 4; Length 1274;
Best Local Similarity 25.3%; Pred. No. 0.0041;
Matches 85; Conservative 26; Mismatches 108; Indels 117; Gaps 17;

QY 86 SSSQHSNTPATQAP-----PLEAPQQA--POPPVAPITTSQPEGLPAQAP-----NTH 132
DB 577 SSGIPTGFAPRIGPQPQPHPSQAFQPQPQPLQLPHLFPQAPGLPPQSPY 636
QY 133 DNAASNLISGRNVDTIINQLEMWGGSDKQVRAALRAAYNPER-----AVEVLY 184
DB 637 PYAPQGVLGQPPPLTQL-----YPGPAQDPLPAHSGALPPFS 676
QY 185 SGIPV-----TAEIATVPIGGQANTTDRAPTGEAGLSGIPNTAPLDLPPOCASNAG 235
DB 677 PGPPQPPHPPLAYGAPSTRPMPQAPLITRPS-SAGQS-----TPSPHLVPSAPSPG 731
QY 236 GG-----AGGQPLDLRNNPQFQAVREMVHTNPQ-----ILQPMLVE 272
DB 732 PGVPVPRPPAARPPCLRRGAAADLLSSPSQ-----HGGTQSPGGQPLLOFTKVD 785
QY 273 LSK-QNPQLRLIEN--HDFLQLLN--EPFEGGEGD-----FLDQPEEDEM 315
DB 786 AAEGRPPQALRIEDPVEHPERLQLOQLEAFRGQLGVGDVLTVMRELQDAQEHAR 845
QY 316 PHAISVTPEEQEAGRLSES-----MGFDRARVI 343
DB 846 GRSI-----AIARCYSLKNRHQDVMFYDSNRVV 873

RESULT 10
US-09-070-060-7
; Sequence 7, Application US/09070060
; Patent No. 5976849
; GENERAL INFORMATION:
; APPLICANT: Rustad, Carolyn M.
; APPLICANT: Ghildyal, Namit
; TITLE OF INVENTION: Human B3 Ubiquitin Protein
; TITLE OF INVENTION: Ligase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZENECA Pharmaceuticals, Inc.
; STREET: 1800 Concord Pike
; CITY: Wilmington
; STATE: DE
; COUNTRY: USA
; ZIP: 19850-5437
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,060

; PRIOR APPLICATION DATA:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/073,839
; FILING DATE: 05-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Higgins, Patrick H
; REGISTRATION NUMBER: 39,709
; REFERENCE/DOCKET NUMBER: PHM.70312
; TELEPHONE: 302.886.4889
; TELEFAX: 302.886.8221
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-09-070-060-7

Query Match 6.6%; Score 126.5; DB 2; Length 156;
Best Local Similarity 35.4%; Pred. No. 0.00078;
Matches 34; Conservative 16; Mismatches 37; Indels 9; Gaps 2;

QY 1 MKLVTTLKGTHTFIRVQPNDTINAVKNIEIQKDSYPWGQQLLIFNGKVLKDESTLE 60
DB 1 MQIFVKILTGTITLEVPSDTIENVK---AKIQKGIIPDQQRLLIFAGKQLEDGRTLS 57
QY 61 ENKYNEDGFLVVML-----SKGKTSSTGTSSSQH 90
DB 58 DYNQKESTLHLVLRGGAKRKKSVTPPKKNH 93

RESULT 11
US-09-051-969A-3
; Sequence 3, Application US/09051969A
; Patent No. 6063598
; GENERAL INFORMATION:
; APPLICANT: ENENKEL, BARBARA
; APPLICANT: GANNON, FRANK
; APPLICANT: BERGMANN, KLAUS
; APPLICANT: NOE, WOLFGANG
; TITLE OF INVENTION: INTENSIVE HOMOLOGOUS PROMOTER OBTAINED
; TITLE OF INVENTION: FROM HAMSTERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/051,969A
; FILING DATE: 1998-09-30
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLESHNER, RAZ E.
; REGISTRATION NUMBER: 34,331
; REFERENCE/DOCKET NUMBER: 0652.1690000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids

```

```
Db      58 DYNIQKESTLHLVLRFGGAKRKXSYTTPQKNKH   93


RESULT 13
US-09-370-746-7
; Sequence 7, Application US/09357746
; Patent No. 6087122
; GENERAL INFORMATION:
; APPLICANT: ZENECA Limited
; TITLE OF INVENTION: HUMAN E3 UBIQUITIN PROTEIN LIGASE
; FILE REFERENCE: PHM.70312.NI
; CURRENT APPLICATION NUMBER: US/09/357,746
; CURRENT FILING DATE: 1999-07-21
; EARLIER APPLICATION NUMBER: US NO. 6087122 60/073,839
; EARLIER FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: US NO. 608712209/070,060
; EARLIER FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-370-746-7

Query Match          6.6%; Score 126.5; DB 3; Length 156;
Best Local Similarity 35.4%; Pred. No. 0.00078;
Matches    34; Conservative    16; Mismatches     37; Indels       9; Gaps           2

Cy      1 MKLTAVKTLKGTHFEIRVPNDTIMAVKKNIETIOGKDSYPMGOQLIFNGKVLDSESTLE 60
        |||:||||::|::|||::|||::|||::|||::|||::|||::|||::|||::|||
Ddb     1 MQIFVKTLTGKTITLEVEPSDTIENVK---AKIQDKEGIPDPQRLLFAGKQLEDGRITLS 57

Cy      61 ENKVNEDGFLVMVL-----SKGTSGSTGTSSSOH 90
        : :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Ddb     58 DYNIQKESTLHLVLRGGAKRKXSYTTPQKNKH   93


RESULT 14
US-09-370-838-205
; Sequence 205, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadon
; APPLICANT: Secretist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FIELD OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 205
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-205

Query Match          6.6%; Score 126.5; DB 4; Length 160;
Best Local Similarity 38.6%; Pred. No. 0.00081;
Matches    39; Conservative    16; Mismatches     35; Indels      11; Gaps            4

Cy      1 MKLTAVKTLKGTHFEIRVPNDTIMAVKKNIETIOGKDSYPMGOQLIFNGKVLDSESTLE 60
        |||:||||::|::|||::|||::|||::|||::|||::|||::|||::|||::|||
Ddb     1 MQIFVKTLTGKTITLEVEPSDTIENVK---AKIQDKEGIPDPQRLLFAGKQLEDGRITLS 57

Cy      61 ENKVNEDG--FLVVMLSKG-----KT-SGSTGTSSSHSNT 93
        : :: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Ddb     58 DYNIQKESTLHLVLRGGWQIFVKTLTGKTITLEVEPSTD 98
```

Search completed: December 17, 2003, 06:21:22  
Job time : 14.3299 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:27:31 ; Search time 18.3376 Seconds  
(without alignments)  
2123.954 Million cell updates/sec

Title: US-09-805-550-2

Perfect score: 405

Sequence: 1 MKNLVKLTGTFEASPD.....BELTANYLLDHGHEFDQQQ 405

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 9616882 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

PIR\_76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	4.4	113	2 H86296	F309.1 protein - A
2	17	4.2	382	2 T14336	RAD23 protein, iso
3	14	3.5	392	2 T04150	RAD23 protein homo
4	13	3.2	246	2 G86296	T24D18.27 protein
5	13	3.2	367	2 F96827	protein F20B17.8 [
6	13	3.2	379	2 T14337	RAD23 protein, iso
7	9	2.2	3069	2 H70656	fatty-acid synthas
8	8	2.0	61	2 B61384	Cowper's gland muc
9	8	2.0	254	1 BVECA	pyruvate dehydroge
10	8	2.0	254	2 E85494	pyruvate dehydroge
11	8	2.0	254	2 E90643	pyruvate dehydroge
12	8	2.0	254	2 AF0415	pyruvate dehydroge
13	8	2.0	254	2 AF0521	pyruvate dehydroge
14	8	2.0	256	2 E82079	pyruvate dehydroge
15	8	2.0	266	2 B32034	streptomycin resis
16	8	2.0	273	2 T03254	probable carbonate
17	8	2.0	289	2 A82006	hypothetical prote
18	8	2.0	300	2 T08106	2'-hydroxyisoflavo
19	8	2.0	339	2 T34406	hypothetical prote
20	8	2.0	393	2 S59499	cellulase egli - s
21	8	2.0	462	2 A84327	photolase/cryptoc
22	8	2.0	566	2 S74633	high affinity sulf
23	8	2.0	643	2 D70697	probable membran
24	8	2.0	653	2 A13404	transposase BME112
25	8	2.0	687	2 D87270	general secretion
26	8	2.0	1235	1 DUBEV1	DNA-directed DNA p
27	8	2.0	1235	1 DUBEH7	DNA-directed DNA p
28	8	2.0	2481	2 D90011	FmtB protein [impo
29	7	1.7	58	2 C84025	hypothetical prote

30	7	1.7	87	2	AG1930	hypothetical prote
31	7	1.7	100	2	B89979	conserved hypothet
32	7	1.7	108	2	AG0699	probable pathogeni
33	7	1.7	128	2	G91250	hypothetical prote
34	7	1.7	129	2	T36732	hypothetical prote
35	7	1.7	139	2	B86810	transcription regu
36	7	1.7	144	2	S53430	adenyde dehydroge
37	7	1.7	144	2	AB2443	hypothetical prote
38	7	1.7	154	2	D86749	riboflavin synthas
39	7	1.7	163	2	C90281	hypothetical prote
40	7	1.7	176	2	F75363	hypothetical prote
41	7	1.7	176	2	E95274	hypothetical prote
42	7	1.7	183	2	F71304	probable ribosome
43	7	1.7	188	2	I51347	MHC class I - Atla
44	7	1.7	218	2	H84364	membrane protein [
45	7	1.7	226	2	T29210	hypothetical prote
46	7	1.7	227	2	T45388	hypothetical prote
47	7	1.7	228	2	S46955	microfilarial shea
48	7	1.7	231	2	E86338	protein F5M15.16 [
49	7	1.7	231	2	A13283	transcription regu
50	7	1.7	232	2	AF3406	hypothetical expor
51	7	1.7	238	2	H83559	hypothetical prote
52	7	1.7	242	2	E70822	hypothetical prote
53	7	1.7	245	2	AD0329	conserved hypothet
54	7	1.7	245	2	AC0830	conserved hypothet
55	7	1.7	247	2	T50874	hypothetical cytoc
56	7	1.7	252	2	JC5237	osmotin-like prote
57	7	1.7	252	2	T29527	hypothetical prote
58	7	1.7	254	2	C82868	plasmid replicatio
59	7	1.7	254	2	T25073	hypothetical prote
60	7	1.7	256	2	S76756	1-(5-phosphoribosy
61	7	1.7	256	2	B32393	T-cell antigen 4-1
62	7	1.7	257	2	T46548	probable transcrip
63	7	1.7	258	2	A75618	probable chromosom
64	7	1.7	264	2	G83245	probable ATP-bindi
65	7	1.7	266	2	T10644	hypothetical prote
66	7	1.7	280	2	S72844	beta-aspartate met
67	7	1.7	280	2	A70513	hypothetical prote
68	7	1.7	280	2	A61047	ectodermal (ect) -
69	7	1.7	281	2	C87073	conserved hypothet
70	7	1.7	281	2	A43749	lignin beta-ether
71	7	1.7	290	2	AG3297	biopolymer transpo
72	7	1.7	292	2	T18584	F5C11.2 - Caenorh
73	7	1.7	301	2	C96961	cell division prot
74	7	1.7	303	2	F81318	probable galactosy
75	7	1.7	303	2	F87629	hypothetical prote
76	7	1.7	307	2	C75584	hypothetical prote
77	7	1.7	315	2	T03733	hypothetical prote
78	7	1.7	322	2	F88029	protein F46F5.5 [i
79	7	1.7	324	2	T04478	probable carbonate
80	7	1.7	330	2	G84137	phosphotransacetyl
81	7	1.7	335	2	A84170	glyceroldehyde 3-p
82	7	1.7	338	2	H87510	polyphenyl synthet
83	7	1.7	338	2	S42594	hypothetical prote
84	7	1.7	340	2	A82218	transcription regu
85	7	1.7	343	2	I51348	MHC class I - Atla
86	7	1.7	347	2	T31755	hypothetical prote
87	7	1.7	351	2	T33976	hypothetical prote
88	7	1.7	358	2	S44233	strP protein - Str
89	7	1.7	364	1	BVEIC	incC protein - pla
90	7	1.7	365	2	I84432	MHC class I protei
91	7	1.7	366	2	JU0469	type II site-speci
92	7	1.7	378	2	H82640	conserved hypothet
93	7	1.7	385	2	G71467	probable succinyl-
94	7	1.7	387	2	T30241	hypothetical prote
95	7	1.7	393	2	E82283	conserved hypothet
96	7	1.7	396	2	A82532	translation elonga
97	7	1.7	400	1	QREBVT	glycine betaine/pr
98	7	1.7	400	1	BVECPV	glycine betaine/l-
99	7	1.7	400	2	D91071	hypothetical prote
100	7	1.7	400	2	G85915	hypothetical prote

## ALIGNMENTS

```

RESULT 1
H86296
F30a.1 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C:Accession: H86296
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziah
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H86296
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-113 <STO>
A:Cross-references: GB:AE005172; NID:g4966345; PIDN:RAD34676.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match      4.4%; Score 18; DB 2; Length 113;
Best Local Similarity 100.0%; Pred. No. 8.4e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 NPQILQMLQELGKQNPQ 312
Db 5 NPQILQMLQELGKQNPQ 22

RESULT 2
T14336
RAD23 protein, isoform I - carrot
C:Species: Daucus carota (carrot)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: T14336
R:Sturm, A.; Leinhardt, S.
Plant J. 13, 815-821, 1998
A:Title: Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in yeast.
A:Reference number: Z17989; MUID:98345997; PMID:9681019
A:Accession: T14336
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-382 <STU>
A:Cross-references: EMBL:Y12013; NID:g1914682; PIDN:CAA72741.1; PID:g1914683
A:Experimental source: subspecies Queen Anne's Lace, isolate W001C
C:Genetics:
A:Gene: RAD23-1
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match      4.2%; Score 17; DB 2; Length 382;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 NPQILQMLQELGKQNPQ 311
Db 282 NPQILQMLQELGKQNPQ 298

RESULT 3
T04150
RAD23 protein homolog - rice
C:Species: Oryza sativa (rice)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: T04150
R:Schultz, T.F.; Quatrano, R.S.

```

```

Plant Mol. Biol. 34, 557-562, 1997
A>Title: Characterization and expression of a rice RAD23 gene.
A:Reference number: Z08695; MUID:97369378; PMID:9225866
A:Accession: T04150
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-392 <SCH>
A:Cross-references: EMBL:U63530; NID:g1488296; PIDN:AAB65841.1; PID:g1488297
A:Experimental source: cv. Nipponbare
C:Genetics:
A:Gene: RAD23
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match      3.5%; Score 14; DB 2; Length 392;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 VADVKRIIETTQGG 36
Db 23 VADVKRIIETTQGG 36

RESULT 4
G86296
T24D18.27 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C:Accession: G86296
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziah
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallor
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86296
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <STO>
A:Cross-references: GB:AE005172; NID:g6587822; PIDN:AAF18513.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match      3.2%; Score 13; DB 2; Length 246;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 RALRAAYNNPERA 194
Db 155 RALRAAYNNPERA 167

RESULT 5
F96827
protein F20B17.8 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
C:Accession: F96827
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziah
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallor
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712

```

A:Accession: F96827  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-367 <STO>  
A:Cross-references: GB:AE005173; NID:g7715605; PIDN:AAF68123.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F20B17.8  
A:Map position: 1  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 3.2%; Score 13; DB 2; Length 367;  
Best Local Similarity 100.0%; Pred. No. 0.0002;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 RALRAAYNNPERA 194  
|||||  
DB 168 RALRAAYNNPERA 180  
|||||

RESULT 6  
T14337  
RAD23 protein, isoform II - carrot  
C:Species: Daucus carota (carrot)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: T14337  
R:Sturm, A.; Leinhardt, S.  
Plant J. 13, 815-821, 1998  
A:Title: Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in yeast.  
A:Reference number: Z17989; MUID:98345997; PMID:9681019  
A:Accession: T14337  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-379 <STU>  
A:Cross-references: EMBL:Y12014; NID:g1914684; PIDN:CAAT72742.1; PID:g1914685  
A:Experimental source: subspecies Queen Anne's Lace, isolate W001C  
C:Genetics:  
A:Gene: RAD23-2  
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 3.2%; Score 13; DB 2; Length 379;  
Best Local Similarity 100.0%; Pred. No. 0.00021;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 RALRAAYNNPERA 194  
|||||  
DB 172 RALRAAYNNPERA 184  
|||||

RESULT 7  
H70656  
fatty-acid synthase (EC 2.3.1.85) - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70656  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: H70656  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-3069 <COL>  
A:Cross-references: GB:Z83863; GB:AL123456; NID:g3261685; PIDN:CAB06201.1; PID:g3261686  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: fas  
C:Superfamily: Brevibacterium ammoniagenes fatty-acid synthase  
C:Keywords: acyltransferase; coenzyme A

Query Match 2.2%; Score 9; DB 2; Length 3069;

Best Local Similarity 100.0%; Pred. No. 12;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 AAPVAPAAAS 108  
|||||  
DB 1754 AAPVAPAAAS 1762  
|||||

RESULT 8  
B61384  
Cowper's gland mucin glycoprotein - pig (fragments)  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 07-Feb-1997  
C:Accession: B61384  
R:Sangadala, S.; Kim, D.; Brewer, J.M.; Mendicino, J.  
Mol. Cell. Biochem. 102, 71-93, 1991  
A:Title: Subunit structure of deglycosylated human and swine trachea and Cowper's gland  
A:Reference number: A61384; MUID:91270244; PMID:2052001  
A:Accession: B61384  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-61 <SAN>  
C:Keywords: glycoprotein

Query Match 2.0%; Score 8; DB 2; Length 61;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 PSPASAGP 246  
|||||  
DB 12 PSPASAGP 19  
|||||

RESULT 9  
BV6CA  
pyruvate dehydrogenase complex repressor - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 31-Mar-1989 #sequence\_revision 05-Dec-1997 #text\_change 01-Mar-2002  
C:Accession: A64734; A30263; S45192; I53482  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; .A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: A64734  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-254 <BLAT>  
A:Cross-references: GB:AE000120; GB:U00096; NID:g1786298; PIDN:AAC73224.1; PID:g178630  
A:Experimental source: strain K-12, substrain MGL655  
R:Stephens, P.E.; Darlison, M.G.; Lewis, H.M.; Guest, J.R.  
Eur. J. Biochem. 133, 153-162, 1983  
A:Title: The pyruvate dehydrogenase complex of Escherichia coli K12. Nucleotide sequen  
A:Reference number: A91130; MUID:83209630; PMID:6343085  
A:Accession: A30263  
A:Molecule type: DNA  
A:Residues: 1-229, 'DFARQK', 236 <STE>  
A:Cross-references: EMBL:V01498  
A:Note: the gene for this protein, called gene A by the authors, is situated upstream  
R:Fujita, N.  
submitted to the EMBL Data Library, January 1994  
A:Reference number: S45181  
A:Accession: S45192  
A:Molecule type: DNA  
A:Residues: 1-229, 'DFARQK', 236 <FUJ>  
A:Cross-references: EMBL:D26562; NID:g473770; PIDN:BAA05571.1; PID:g473782  
A:Experimental source: strain K-12 substrain W3110  
R:Haydon, D.J.; Quail, M.A.; Guest, J.R.  
FEBS Lett. 336, 43-47, 1993  
A:Title: A mutation causing constitutive synthesis of the pyruvate dehydrogenase compl.  
A:Reference number: I53482; MUID:94085588; PMID:8262214  
A:Accession: I53482  
A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA  
A:Residues: 1-117, 'C', 119-254 <RES>  
A:Cross-references: GB:S67363; NID:9455921; PIDN:AAB29356.1; PID:9455922  
C:Genetics:

A:Gene: pdhR

C:Map position: 3 min  
C:Superfamily: pyruvate dehydrogenase complex repressor  
C:Keywords: DNA binding, transcription regulation

Query Match 2.0%; Score 8; DB 1; Length 254;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRL 366

DB 52 REAIQRL 59

#### RESULT 10

E85494 Pyruvate dehydrogenase complex repressor - Escherichia coli (strain O157:H7, substrain E

C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 19-Jul-2002

C:Accession: E85494  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: E85494

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-254 <STO>

A:Cross-references: GB:AE005174; NID:912512822; PIDN:AAG54417.1; GSPDB:GN00145; UMGF:Z01

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: pdhR

C:Superfamily: pyruvate dehydrogenase complex repressor

Query Match 2.0%; Score 8; DB 2; Length 254;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRL 366

DB 52 REAIQRL 59

#### RESULT 11

E90643 Pyruvate dehydrogenase complex repressor - Escherichia coli (strain O157:H7, substrain R

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 19-Jul-2002

C:Accession: E90643

R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: E90643

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-254 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA833540.1; PID:gl3359573; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs0117

C:Superfamily: pyruvate dehydrogenase complex repressor

Query Match 2.0%; Score 8; DB 2; Length 254;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRL 366

DB 52 REAIQRL 59

#### RESULT 12

AF0415

Pyruvate dehydrogenase complex repressor [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 19-Jul-2002

C:Accession: AF0415

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AF0415

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-254 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC92650.1; PID:gl5981347; GSPDB:GN00175

C:Genetics:

A:Gene: pdhR

C:Superfamily: pyruvate dehydrogenase complex repressor

Query Match 2.0%; Score 8; DB 2; Length 254;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRL 366

DB 52 REAIQRL 59

#### RESULT 13

AF0521

Pyruvate dehydrogenase complex repressor [imported] - Salmonella enterica subsp. enter

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Notes: This species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AF0521

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, P.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AF0521

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-254 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD01310.1; PID:gl6501438; GSPDB:GN00176

C:Genetics:

A:Gene: STY0174

C:Superfamily: pyruvate dehydrogenase complex repressor

Query Match 2.0%; Score 8; DB 2; Length 254;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRL 366

DB 52 REAIQRL 59

#### RESULT 14

E82079

Pyruvate dehydrogenase complex repressor VC2415 [imported] - Vibrio cholerae (strain

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 19-Jul-2002

C:Accession: E82079



R.Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, J.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A>Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
 A;Reference number: AB2035; MUID:20406833; PMID:10952301  
 A;Accession: E82079  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-256 <HEI>  
 A;Cross-references: GB:AE004311; GB:AE003952; NID:g9656983; PIDN:AAF95558.1; GSPDB:GN001  
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C;Genetics:  
 A;Gene: VC2415  
 A;Map position: 1  
 C;Superfamily: pyruvate dehydrogenase complex repressor

Query Match 2.0%; Score 8; DB 2; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 REAIOQLE 366  
 |||||  
 Db 52 REAIOQLE 59

RESULT 15  
 B23034  
 streptomycin resistance protein - *Klebsiella* sp. transposon Tn5  
 C;Species: *Klebsiella* sp.  
 C;Date: 20-Aug-1987 #sequence\_revision 20-Aug-1987 #text\_change 19-Feb-2000  
 A;Accession: B23034  
 R;Mazodier, P.; Cossart, P.; Gitraud, E.; Gasser, F.  
 Nucleic Acids Res. 13 195-205, 1985  
 A>Title: Completion of the nucleotide sequence of the central region of Tn5 confirms the  
 A;Reference number: A93559; NID:85215465; PMID:3889831  
 A;Accession: B23034  
 A;Molecule type: DNA  
 A;Residues: 1-266 <MAZ>  
 C;Superfamily: streptomycin 6-kinase  
 C;Keywords: antibiotic resistance

Query Match 2.0%; Score 8; DB 2; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 APAASVAR 111  
 |||||  
 Db 132 APAASVAR 139

RESULT 16  
 T03254  
 probable carbonate dehydratase (EC 4.2.1.1), chloroplast - rice  
 N;Alternate names: carbonic anhydrase  
 C;Species: *Oryza sativa* (rice)  
 C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 22-Jun-1999  
 A;Accession: T03254  
 R;Suzuki, S.; Burnell, J.N.  
 submitted to the EMBL Data Library, April 1994  
 A;Description: Rice carbonic anhydrase: purification, location, and sequence analysis of  
 A;Reference number: Z14863  
 A;Accession: T03254  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-273 <SUZ>  
 A;Cross-references: EMBL:U08404; NID:g606816; PIDN:AAA86943.1; PID:g606817  
 A;Experimental source: leaf  
 C;Genetics:  
 A;Genome: nuclear  
 C;Function:  
 A;Description: catalyzes the reversible hydration of carbon dioxide  
 C;Superfamily: *Escherichia coli* carbonate dehydratase

C;Keywords: carbon-oxygen lyase; chloroplast; hydro-lyase

Query Match 2.0%; Score 8; DB 2; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 AAPVAPAA 107  
 |||||  
 Db 64 AAPVAPAA 71

RESULT 17  
 AE2006  
 hypothetical protein alr1603 [imported] - *Nostoc* sp. (strain PCC 7120)  
 C;Species: *Nostoc* sp. PCC 7120  
 A;Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
 C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C;Accession: AE2006  
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguc  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,  
 DNA Res. 8, 205-213, 2001  
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium A  
 A;Reference number: AB1807; MUID:21595285; PMID:11759840  
 A;Accession: AE2006  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-289 <KUR>  
 A;Cross-references: GB:BA000019; PIDN:BA877969.1; PID:gl7135423; GSPDB:GN00179  
 A;Experimental source: strain PCC 7120  
 C;Genetics:  
 A;Gene: alr1603

Query Match 2.0%; Score 8; DB 2; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 AAPAATVA 139  
 |||||  
 Db 56 AAPAATVA 63

RESULT 18  
 T08106  
 2'-hydroxyisoflavone reductase (EC 1.1.1.45) - European white birch  
 C;Species: *Betula pendula* (European white birch)  
 C;Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 26-May-2000  
 R;Schmitz, N.; Frank, E.; Scheurer, S.; Hausteinh, D.; Vieths, S.  
 submitted to the EMBL Data Library, February 1998  
 A;Description: Characterization of a new birch pollen allergen, representing an IPRH P  
 A;Reference number: Z16359  
 A;Accession: T08106  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-300 <SCH>  
 A;Cross-references: EMBL:AF047896; NID:g2935415; PIDN:AAC05116.1; PID:g2935416  
 C;Superfamily: 2'-hydroxyisoflavone reductase  
 C;Keywords: oxidoreductase

Query Match 2.0%; Score 8; DB 2; Length 300;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TNFEIEAS 18  
 |||||  
 Db 276 TNFEIEAS 283

RESULT 19  
 T34406  
 hypothetical protein EGAP9.3 - *Caenorhabditis elegans*  
 C;Species: *Caenorhabditis elegans*  
 C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T34406  
R:Geisel, C.  
submitted to the EMBL Data Library, November 1996  
A:Description: The sequence of *C. elegans* cosmid EGAP9.  
A:Reference number: Z21519  
A:Accession: T34406  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-339 <GEI>  
A:Cross-references: EMBL:U80026; PIDN:RAC25845.1; GSPDB:GN00023; CESP:EGAP9.3  
A:Experimental source: strain Bristol N2; clone EGAP9  
C:Genetics:  
A:Gene: CESP:EGAP9.3  
A:Map position: 5  
A:Introns: 40/3; 78/3; 117/2; 152/1; 208/1; 238/3; 267/2; 309/2

Query Match 2.0%; Score 8; DB 2; Length 339;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 RAIDLYLS 200  
|||||  
Db 143 RAIDLYLS 150

RESULT 20  
S59499  
cellulase egII - smut fungus (*Ustilago maydis*)  
C:Species: *Ustilago maydis* (corn smut)  
C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 28-May-1999  
C:Accession: S59499  
R:Schauwecker, F.; Wanner, G.; Kahmann, R.  
Biol. Chem. Hoppe-Seyler 376, 617-625, 1995  
A:Title: Filament-specific expression of a cellulase gene in the dimorphic fungus *Ustilago maydis*  
A:Reference number: S59499; MUID:96145728; PMID:8590631  
A:Accession: S59499  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-393 <SCH>  
A:Cross-references: GB:S81598; NID:gl478378; PIDN:AAB36147.1; PID:gl478379

Query Match 2.0%; Score 8; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 PSQGSNPG 266  
|||||  
Db 362 PSQGSNPG 369

RESULT 21  
A84327  
photolase/cryptochrome [imported] - *Halobacterium* sp. NRC-1  
C:Species: *Halobacterium* sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: A84327  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo, J.; Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of *Halobacterium* species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: A84327  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-462 <STO>  
A:Cross-references: GB:AE004437; NID:gl0581208; PIDN:AAG19981.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: phr1  
C:Superfamily: deoxyribodipyrimidine photo-lyase

Query Match 2.0%; Score 8; DB 2; Length 462;

Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 VRALRAAY 188  
|||||  
Db 53 VRALRAAY 60

RESULT 22  
S74633  
high affinity sulfate transporter - *Synechocystis* sp. (strain PCC 6803)  
A:Alternate names: hypothetical protein slr1776  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S74633  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S74633  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-566 <KAN>  
A:Cross-references: EMBL:D90900; GB:AB001339; NID:gl651768; PIDN:BAAL6785.1; PID:gl651  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Start codon: GTG  
C:Superfamily: integral membrane protein HP0228

Query Match 2.0%; Score 8; DB 2; Length 566;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 GVPSSGSN 264  
|||||  
Db 155 GVPSSGSN 162

RESULT 23  
D70697  
probable membraneprotein - *Mycobacterium tuberculosis* (strain H37RV)  
C:Species: *Mycobacterium tuberculosis*  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: D70697  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, R.; Andream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
C:Accession: D70697  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-643 <COL>  
A:Cross-references: GB:Z80343; GB:AL123456; NID:gl3261648; PIDN:CAB02471.1; PID:e267405  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv3792

Query Match 2.0%; Score 8; DB 2; Length 643;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 LQQLAAAV 348  
|||||  
Db 29 LQQLAAAV 36

RESULT 24

```

AI3404
transposase BMEI1223 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AI3404
R:DeiVecchio, V.G.; Kaparal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzeck, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AI3404
A:Status: preliminary
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-653 <KUR>
A:Cross-references: GB:AB008917; PIDN:AAL52404.1; PID:G17983205; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI1223
A:Map position: 1

Query Match 2.0%; Score 8; DB 2; Length 653;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 SGALDALR 280
|||||
DB 255 SGALDALR 262

RESULT 25
DB87270
general secretion pathway protein D [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: DB87270
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.R.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.B.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: DB87270
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-687 <STO>
A:Cross-references: GB:AE005673; NID:G13421292; PIDN:AAK22160.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0173

Query Match 2.0%; Score 8; DB 2; Length 687;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 GASTATTA 90
|||||
DB 307 GASTATTA 314

RESULT 26
DBEVI
DNA-directed DNA polymerase (EC 2.7.7.7) - human herpesvirus 1
C:Species: human herpesvirus 1
A:Note: host Homo sapiens (man)
C:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 28-May-1999
C:Accession: A00715
R:Quinn, J.P.; McGeoch, D.J.
Nucleic Acids Res. 13, 8143-8163, 1985
A:Title: DNA sequence of the region in the genome of herpes simplex virus type 1 contain
A:Reference number: A93601; MUID:86067223; PMID:2999714
A:Accession: A00715
A:Molecule type: DNA
A:Residues: 1-1235 <QUI>
A:Cross-references: GB:X03181; GB:M12356; NID:G59862; PIDN:CAA26941.1; PID:G59864

```

```

A:Experimental source: strain 17
C:Genetics:
A:Map position: 0.41-0.43
C:Superfamily: herpesvirus DNA-directed DNA polymerase
C:Keywords: DNA binding; DNA replication; nucleotidyltransferase

Query Match 2.0%; Score 8; DB 1; Length 1235;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 TLAQPAAP 102
|||||
DB 325 TLAQPAAP 332

RESULT 27
DJBEH7
DNA-directed DNA polymerase (EC 2.7.7.7) - human herpesvirus 1 (strain 17)
C:Species: human herpesvirus 1
C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000
C:Accession: C30085
R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Pe
J. Gen. Virol. 69, 1531-1574, 1988
A:Title: The complete DNA sequence of the long unique region in the genome of herpes s
A:Reference number: A30083; MUID:88274327; PMID:2839594
A:Accession: C30085
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1235 <MCG>
A:Cross-references: GB:X14112; GB:D00317; GB:D00374; GB:S40593; NID:G1944536; PIDN:CAA
C:Genetics:
A:Gene: UL30
C:Superfamily: herpesvirus DNA-directed DNA polymerase
C:Keywords: DNA binding; DNA replication; nucleotidyltransferase

Query Match 2.0%; Score 8; DB 1; Length 1235;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 TLAQPAAP 102
|||||
DB 325 TLAQPAAP 332

RESULT 28
D90011
FmtB protein [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: D90011
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Og
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D90011
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2481 <KUR>
A:Cross-references: GB:BA000018; PID:G13701961; PIDN:BA843253.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: fmtB(mrp)

Query Match 2.0%; Score 8; DB 2; Length 2481;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQAAPVA 104
|||||
DB 144 AQAAPVA 151

```

```

RESULT 29
C84025
hypothetical protein BH3003 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C84025
R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A: Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A: Reference number: A83650; MUID:20512582; PMID:11058132
A: Accession: C84025
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-58 <STO>
A: Cross-references: GB:AP001517; GB:BA000004; NID:G10175500; PIDN:BAB06722.1; GSPDB:GN00
A: Experimental source: strain C-125
C: Genetics:
A: Gene: BH3003

Query Match 1.7%; Score 7; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 EREAIQ 363
Db 10 EREAIQ 16

RESULT 30
AG1930
hypothetical protein asl0994 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A: Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C: Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C: Accession: AG1930
R: Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A: Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A: Reference number: AB1807; MUID:21595285; PMID:11759840
A: Accession: AG1930
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-87 <KUR>
A: Cross-references: GB:BA000019; PIDN:BAB72951.1; PID:G17130340; GSPDB:GN00179
A: Experimental source: strain PCC 7120
C: Genetics:
A: Gene: asl0994

Query Match 1.7%; Score 7; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 LRLVNES 331
Db 5 LRLVNES 11

RESULT 31
B89979
conserved hypothetical protein SA1723 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C: Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C: Accession: B89979
R: Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hitamatsu, K.
Lancet 357, 1225-1240, 2001
A: Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A: Reference number: A89758; MUID:21311952; PMID:11418146
A: Accession: B89979
A: Status: preliminary

```

```

A: Molecule type: DNA
A: Residues: 1-100 <KUR>
A: Cross-references: GB:BA000018; PID:G13701700; PIDN:BAB42993.1; GSPDB:GN00149
A: Experimental source: strain N315
C: Genetics:
A: Gene: SA1723
C: Superfamily: Bacillus subtilis hypothetical protein yerc

Query Match 1.7%; Score 7; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 SGASTAT 88
Db 66 SGASTAT 72

RESULT 32
AG0699
probable pathogenicity island protein sseA [imported] - Salmonella enterica subsp. ent
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A: Note: this species has also been called Salmonella typhi
C: Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C: Accession: AG0699
R: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farra
. S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A: Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K
A: Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser
A: Reference number: AB0502; MUID:21534947; PMID:11677608
A: Accession: AG0699
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-108 <PAR>
A: Cross-references: GB:AL513382; PIDN:CAD01968.1; PID:G16502810; GSPDB:GN00176
C: Genetics:
A: Gene: sseA

Query Match 1.7%; Score 7; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 NLFPOGV 258
Db 59 NLFPOGV 65

RESULT 33
G91250
hypothetical protein ECs4975 [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C: Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C: Accession: G91250
R: Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A: Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ge
A: Reference number: A99629; MUID:21156231; PMID:11258796
A: Accession: G91250
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-128 <HAY>
A: Cross-references: GB:BA000007; PIDN:BAB39398.1; PID:G13364451; GSPDB:GN00154
A: Experimental source: strain O157:H7, substrain RIMD 0509952
C: Genetics:
A: Gene: ECs4975

Query Match 1.7%; Score 7; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 TVAATDD 143

```

Db 6 TAAATDD 12  
 |||||  
 RESULT 34  
 T36732  
 hypothetical protein SCH69.33 - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T36732  
 R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, June 1999  
 A:Reference number: Z21612  
 A:Accession: T36732  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-129 <MUR>  
 A:Cross-references: EMBL:AL079308; PIDN:CAB45230.1; GSPDB:GN00070; SCORDB:SCH69.33  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCORDB:SCH69.33

Query Match 1.7%; Score 7; DB 2; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 345 AAAPVPT 351  
 |||||  
 Db 32 AAAPVPT 38

RESULT 35  
 B86810  
 transcription regulator [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
 C:Species: Lactococcus lactis subsp. lactis  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: B86810  
 R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich  
 Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
 A:Reference number: A86625; MUID:21235186; PMID:11337471  
 A:Accession: B86810  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-139 <STO>  
 A:Cross-references: GB:AE005176; PID:gl2724476; PIDN:AAK05580.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: imaE

Query Match 1.7%; Score 7; DB 2; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 321 QAEFLRL 327  
 |||||  
 Db 125 QAEFLRL 131

RESULT 36  
 S53430  
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3), cytosolic - rat (fragments)  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 01-Aug-1995 #sequence\_revision 03-Nov-1995 #text\_change 03-Jun-2002  
 C:Accession: S53430  
 R:Labrecque, J.; Dumas, F.; Lacroix, A.; Bhat, P.V.  
 Biochem. J. 305, 681-684, 1995  
 A:Title: A novel isoenzyme of aldehyde dehydrogenase specifically involved in the biosyn  
 A:Reference number: S53430; MUID:95134257; PMID:7832787  
 A:Accession: S53430  
 A:Molecule type: protein  
 A:Residues: 1-18;19-30;31-56;57-77;78-99;100-144 <LAB>  
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: blocked amino end; cytosol; NAD; oxidoreductase

Query Match 1.7%; Score 7; DB 2; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 227 SPAQPAV 233  
 |||||  
 Db 2 SPAQPAV 8

RESULT 37  
 AB2443  
 hypothetical protein all5098 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AB2443  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguc  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium A  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AB2443  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-144 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA876797.1; PID:gl7134236; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all5098

Query Match 1.7%; Score 7; DB 2; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 VATAETA 124  
 |||||  
 Db 24 VATAETA 30

RESULT 38  
 D86749  
 riboflavin synthase (EC 2.5.1.9) beta chain [imported] - Lactococcus lactis subsp. lac  
 C:Species: Lactococcus lactis subsp. lactis  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: D86749  
 R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehr  
 Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
 A:Reference number: A86625; MUID:21235186; PMID:11337471  
 A:Accession: D86749  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-154 <STO>  
 A:Cross-references: GB:AE005176; PID:gl2723940; PIDN:AAK05094.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: ribH  
 C:Superfamily: riboflavin synthase beta chain  
 C:Keywords: transferase

Query Match 1.7%; Score 7; DB 2; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 273 SGALDAL 279  
 |||||  
 Db 32 SGALDAL 38

RESULT 39  
 C90281

hypothetical protein SS01269 [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
 C:Accession: C90281  
 R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
 arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: C90281  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-163 <KUR>  
 A:Cross-references: GB:AE006641; NID:g13814464; PIDN:AAK41506.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: SS01269

Query Match 1.7%; Score 7; DB 2; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 DVKRIIE 31  
 |||||  
 DB 101 DVKRIIE 107

RESULT 40  
 F75363  
 hypothetical protein - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 26-May-2000  
 C:Accession: F75363  
 R;White, O.; Eisen, J.A.; Heideberg, J.P.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: F75363  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-176 <WHI>  
 A:Cross-references: GB:AE002012; GB:AE000513; NID:g6459473; PIDN:AAF11269.1; PID:g645948  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR1708  
 A:Map position: 1  
 C:Superfamily: Deinococcus radiodurans hypothetical protein DR1708

Query Match 1.7%; Score 7; DB 2; Length 176;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 QAAPAT 137  
 |||||  
 DB 27 QAAPAT 33

RESULT 41  
 E95274  
 hypothetical protein Sma0191 [imported] - Sinorhizobium meliloti (strain 1021) magapla  
 C:Species: Sinorhizobium meliloti  
 C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: E95274  
 R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bow  
 ; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9893-9898, 2001  
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilo  
 A:Reference number: A95262; MUID:21396509; PMID:11481432  
 A:Accession: E95274  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-176 <KUR>  
 A:Cross-references: GB:AE006469; PIDN:AAK64759.1; PID:g14523165; GSPDB:GN00165  
 A:Experimental source: strain 1021, megaplasmid pSymA  
 R;Galibert, F.; Finan, T.M.; Long, S.R.; Fuhler, A.; Abola, P.; Ampe, F.; Barloy-Huble  
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Komp, C.; Lelau  
 hehault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,  
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: Sma0191  
 A:Genome: plasmid

Query Match 1.7%; Score 7; DB 2; Length 176;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 AAPAATV 138  
 |||||  
 DB 18 AAPAATV 24

RESULT 42  
 F71304  
 probable ribosome recycling factor - syphilis spirochete  
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
 C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 26-Aug-1999  
 C:Accession: F71304  
 R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gw  
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; Mc  
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
 A:Reference number: A71250; MUID:98332770; PMID:9665876  
 A:Accession: F71304  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-183 <COL>  
 A:Cross-references: GB:AE001235; GB:AE000520; NID:g3322893; PIDN:AAC65577.1; PID:g332  
 A:Experimental source: strain Nicholas  
 C:Genetics:  
 A:Gene: TP0604  
 C:Superfamily: ribosome releasing factor

Query Match 1.7%; Score 7; DB 2; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 DASVADV 26  
 |||||  
 DB 163 DASVADV 169

RESULT 43  
 IS1347  
 MHC class I - Atlantic salmon (fragment)  
 C:Species: Salmo salar (Atlantic salmon)  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 03-Dec-1999  
 C:Accession: IS1347  
 R;Grimholt, U.; Hordvik, I.; Posse, V.M.; Olsaker, I.; Endresen, C.; Lie, O.  
 Immunogenetics 37, 469-473, 1993  
 A:Title: Molecular cloning of major histocompatibility complex class I cDNAs in Atlan  
 A:Reference number: IS1347; MUID:93170896; PMID:8436423  
 A:Accession: IS1347  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-188 <GRI>  
 A:Cross-references: GB:L07605; NID:g213820; PIDN:AAA49602.1; PID:g213821  
 C:Genetics:  
 A:Gene: MHC-I

C;Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 1.7%; Score 7; DB 2; Length 188;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 LVTVPPE 358  
Db 106 LVTVPPE 112

# RESULT 44

H84364  
membrane protein [imported] - Halobacterium sp. NRC-1  
C;Species: Halobacterium sp. NRC-1  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: H84364  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablonski, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.  
A;Reference number: A84160; MUID:20504483; PMID:11016950  
A;Accession: H84364  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-218 <STO>  
A;Cross-references: GB:AE004437; NID:gl0581563; PIDN:AAG20284.1; GSPDB:GN00138  
C;Genetics:  
A;Gene: hmp

Query Match 1.7%; Score 7; DB 2; Length 218;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 VARAPAA 217  
Db 210 VARAPAA 216

# RESULT 45

T29210  
hypothetical protein T20F5.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T29210  
R;Du, Z.; Le, T.T.  
submitted to the EMBL Data Library, November 1996  
A;Description: The sequence of C. elegans coamid T20F5.  
A;Reference number: Z20588  
A;Accession: T29210  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-226 <DUZ>  
A;Cross-references: EMBL:U80442; PIDN:AAB37667.1; GSPDB:GN00019; CESP:T20F5.5  
A;Experimental source: strain Bristol N2; clone T20F5  
C;Genetics:  
A;Gene: CESP:T20F5.5  
A;Map position: 1  
A;Introns: 14/3

Query Match 1.7%; Score 7; DB 2; Length 226;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 PAAPVAP 105  
Db 125 PAAPVAP 131

# RESULT 46

T45388

hypothetical protein [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae  
C;Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 31-Jan-2000  
C;Accession: T45388  
R;Cole, S.T.; Flesselles, B.; Honore, N.  
submitted to the EMBL Data Library, August 1997  
A;Reference number: Z22966  
A;Accession: T45388  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-227 <COL>

A;Cross-references: EMBL:Z98756; PIDN:CABL1458.1  
A;Experimental source: cosmid B2492  
C;Genetics:  
A;Note: MLCB2492.26

Query Match 1.7%; Score 7; DB 2; Length 227;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 VARAPAA 217  
Db 212 VARAPAA 218

# RESULT 47

S46965  
microfilarial sheath protein - nematode (Brugia pahangi)  
C;Species: Brugia pahangi  
C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 19-May-2000  
C;Accession: S46965  
R;Hirzmann, J.; Schnauffer, A.; Hintz, M.; Conratha, F.J.; Stirm, S.; Zahner, H.; Hobo  
submitted to the EMBL Data Library, July 1994  
A;Description: Brugia spp. and Litomosoides carinii: characterization of the shp2 gene  
A;Reference number: S46964  
A;Accession: S46965  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-228 <HIR>  
A;Cross-references: EMBL:Z35445; NID:gs16354; PID:gs16355  
C;Genetics:  
A;Introns: 26/1  
C;Superfamily: hydroxyproline-rich glycoprotein

Query Match 1.7%; Score 7; DB 2; Length 228;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 PAAPVAP 105  
Db 81 PAAPVAP 87

# RESULT 48

E86338  
protein F5M15.16 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: E86338  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: E86338  
A;Status: preliminary  
A;Molecule type: DNA

```

A:Residues: 1-231 <STO>
A:Cross-references: GB:AE005172; NID:g9778602; PIDN:AAF79610.1; GSPDB:GN00141
C:Genetics:
A:Gene: F5M15.16
A:Map position: 1

Query Match      1.7%; Score 7; DB 2; Length 231;
Best Local Similarity 100.0%; Pred. No. 1.e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 ANYLLDH 396
DB 16 ANYLLDH 22

RESULT 49
A:3283
transcription regulator, gntr family BMEI0254 [imported] - Brucella melitensis (strain 1
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: A13283
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letese
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: A13283
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-231 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL52417.1; PID:g17983220; GSPDB:GN00190
A:Gene: BMEI0254
C:Genetics:
A:Experimental source: strain 16M
A:Map position: 1

Query Match      1.7%; Score 7; DB 2; Length 231;
Best Local Similarity 100.0%; Pred. No. 1.e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRL 365
DB 70 REAIQRL 76

RESULT 50
AF3406
hypothetical exported proline-rich protein BMEI1236 [imported] - Brucella melitensis (st
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AF3406
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letese
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3406
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-232 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL52417.1; PID:g17983220; GSPDB:GN00190
A:Gene: BMEI1236
C:Genetics:
A:Experimental source: strain 16M
A:Map position: 1

Query Match      1.7%; Score 7; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 1.e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 PAAPVAP 105
DB 75 PAAPVAP 81

```

```

RESULT 51
H83559
hypothetical protein PA0674 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83559
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Li
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pat
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83559
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-238 <STO>
A:Cross-references: GB:AE004503; GB:AE004091; NID:g9946553; PIDN:AAG04063.1; GSPDB:GNO
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0674

Query Match      1.7%; Score 7; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 PATLAQP 99
DB 207 PATLAQP 213

RESULT 52
E70822
hypothetical protein Rv0730 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70822
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genom
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70822
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-242 <COL>
A:Cross-references: GB:AL021959; GB:AL123456; NID:g3261536; PIDN:CAA17497.1; PID:g2911
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv0730

Query Match      1.7%; Score 7; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 VARAPAA 217
DB 226 VARAPAA 232

RESULT 53
AD0329
conserved hypothetical protein YPO2700 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AD0329
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell
Nature 413, 523-527, 2001

```



A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AD0329  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-245 <KUR>  
A;Cross-references: GB:AL590842; PIDN:CAC92939.1; PID:gl5980678; GSPDB:GN00175  
C;Genetics:  
A;Gene: YPO2700  
C;Superfamily: Bacillus subtilis lactam utilization protein ycsF

Query Match 1.7%; Score 7; DB 2; Length 245;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 QALLQLV 292  
|||||  
Db 17 QALLQLV 23

RESULT 54  
AC0830  
Conserved hypothetical protein STY2835 [imported] - Salmonella enterica subsp. enterica  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AC0830  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, Th. T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AC0830  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-245 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD02791.1; PID:gl6503801; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY2835  
C;Superfamily: hypothetical protein HI0423

Query Match 1.7%; Score 7; DB 2; Length 245;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 VADVKKI 29  
|||||  
Db 42 VADVKKI 48

RESULT 55  
T50874  
Hypothetical cytochrome [imported] - Rubrivivax gelatinosus  
C;Species: Rubrivivax gelatinosus  
C;Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C;Accession: T50874  
R;Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.  
submitted to the EMBL Data Library, November 1999  
A;Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosyn  
A;Reference number: Z25270  
A;Accession: T50874  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-247 <NAG>  
A;Cross-references: EMBL:AB034704; PIDN:BAA94027.1  
A;Experimental source: strain IL144  
C;Genetics:  
A;Note: ORF247

Query Match 1.7%; Score 7; DB 2; Length 247;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 AAPVAPA 106  
|||||  
Db 48 AAPVAPA 54

RESULT 56  
JC5237  
osmotin-like protein precursor - tomato  
C;Species: Lycopersicon esculentum (tomato)  
C;Date: 13-Mar-1997 #sequence\_revision 13-Mar-1997 #text\_change 21-Jul-2000  
C;Accession: JC5237  
R;Chen, R.; Wang, F.; Smith, A.G.  
Gene 179, 301-302, 1996  
A;Title: A flower-specific gene encoding an osmotin-like protein from Lycopersicon esc  
A;Reference number: JC5237; MUID:97128324; PMID:8972917  
A;Contents: flower  
A;Accession: JC5237  
A;Status: preliminary; nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-252 <CHE>  
A;Cross-references: GB:L76632; NID:gl220536; PIDN:AAB41124.1; PID:gl220537  
C;Comment: This protein is rich in cysteine content and play a part of the flowers def  
C;Superfamily: thaumatin I  
F;1-23/Domain: signal sequence #status predicted <SIG>  
F;24-252/Product: osmotin-like protein #status predicted <MAT>

Query Match 1.7%; Score 7; DB 2; Length 252;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 APATLAQ 98  
|||||  
Db 113 APATLAQ 119

RESULT 57  
T29527  
hypothetical protein F27C1.2 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T29527  
R;Wu, X.; Le, T.T.  
submitted to the EMBL Data Library, November 1996  
A;Description: The sequence of C. elegans cosmid F27C1.  
A;Reference number: Z20636  
A;Accession: T29527  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-252 <WUX>  
A;Cross-references: EMBL:U80441; PIDN:AAB37655.1; GSPDB:GN00019; CESP:F27C1.2  
A;Experimental source: strain Bristol N2; clone F27C1  
C;Genetics:  
A;Gene: CESP:F27C1.2  
A;Map position: 1  
A;Introns: 23/3; 100/1; 142/3; 216/1

Query Match 1.7%; Score 7; DB 2; Length 252;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 107 ASVARTP 113  
|||||  
Db 172 ASVARTP 178

RESULT 58  
C82868  
plasmid replication protein XFa0060 [imported] - Xylella fastidiosa (strain 9a5c)  
C;Species: Xylella fastidiosa  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 24-May-2001  
C;Accession: C82868

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: C82868  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-254 <SIM>  
 A:Cross-references: GB:AEO03851; NID:G9112238; PIDN:AAF95628.1; GSPDB:GN00130; XFSC:XFAO  
 A:Experimental source: strain 945C  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H  
 ae-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Content: annotation  
 C:Genetics:  
 A:Gene: XFA0060  
 A:Genome: plasmid  
 A:Note: plasmid pXF5.1  
 C:Superfamily: regulatory protein spo0J

Query Match 1.7%; Score 7; DB 2; Length 254;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 275 ALDALRQ 281  
 Db 192 ALDALRQ 198  
 |||||

RESULT 59  
 T25073  
 hypothetical protein T21C9.9 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T25073  
 R:McMurray, A.  
 submitted to the EMBL Data Library, May 1996  
 A:Reference number: Z19977  
 A:Accession: T25073  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-254 <WIL>  
 A:Cross-references: EMBL:Z73098; PIDN:CAA97336.1; GSPDB:GN00023; CESP:T21C9.9  
 A:Experimental source: clone T21C9  
 C:Genetics:  
 A:Gene: CESP:T21C9.9  
 A:Map position: 5  
 A:Introns: 22/2; 78/3

Query Match 1.7%; Score 7; DB 2; Length 254;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 QPAAPVA 104  
 Db 133 QPAAPVA 139  
 |||||

RESULT 60  
 S76756  
 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide  
 C:Species: *Synechocystis* sp.

A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 03-Jun-2002  
 A:Accession: S76756  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N  
 o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocyst*  
 s.  
 A:Reference number: S74322; MUID:97061201; PMID:8905231  
 A:Accession: S76756  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-256 <KAN>  
 A:Cross-references: EMBL:D90916; GB:AB001339; NID:G1653715; PIDN:BAAL8668.1; PID:G1653  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Superfamily: N-(5'-phospho-D-riboseylformimino)-5-amino-1-(5'-phosphoribosyl)-4-imid  
 C:Keywords: intramolecular oxidoreductase; isomerase

Query Match 1.7%; Score 7; DB 2; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 LDALRQL 282  
 Db 179 LDALRQL 185  
 |||||

RESULT 61  
 B32393  
 T-cell antigen 4-1BB precursor - mouse  
 C:Species: *Mus musculus* (house mouse)  
 C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 21-Jul-2000  
 C:Accession: B32393; I48879  
 R:Kwon, B.S.; Weisman, S.M.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 1963-1967, 1989  
 A:Title: cDNA sequence of two inducible T-cell genes.  
 A:Reference number: A32393; MUID:89184547; PMID:2784565  
 A:Accession: B32393  
 A:Molecule type: mRNA  
 A:Residues: 1-256 <KWO>  
 A:Cross-references: GB:J04492; NID:G201121; PIDN:AAA40167.1; PID:G201122  
 R:Kwon, B.S.; Kozak, C.A.; Kim, K.K.; Pickard, R.T.  
 J. Immunol. 152, 2256-2262, 1994  
 A:Title: Genomic Organization and Chromosomal Localization of the T-Cell Antigen 4-1B  
 A:Reference number: I48879; MUID:94179805; PMID:8133039  
 A:Accession: I48879  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-256 <RES>  
 A:Cross-references: EMBL:U02567; NID:G1117783; PIDN:AAA93113.1; PID:G409178  
 C:Genetics:  
 A:Introns: 35/1; 69/1; 115/1; 138/2; 156/3; 181/1; 225/1  
 C:Superfamily: CD27 antigen; NGF receptor repeat homology  
 C:Keywords: transmembrane protein  
 F1-23/Domain: signal sequence #status predicted <SIG>  
 F1-24-256/Product: 4-1BB protein #status predicted <NAT>

Query Match 1.7%; Score 7; DB 2; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 PEGPGGG 338  
 Db 175 PEGPGGG 181  
 |||||

RESULT 62  
 T46548  
 probable transcription activator [imported] - *Streptomyces violaceoruber*  
 C:Species: *Streptomyces violaceoruber*  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 11-May-2000  
 C:Accession: T46548  
 R:Ichinose, K.; Bedford, D.J.; Tornus, D.; Bechthold, A.; Bibb, M.J.; Revill, W.P.; F

Chem. Biol. 5, 647-659, 1998

A;Title: The granaticin biosynthetic gene cluster of Streptomyces violaceoruber Tu22: se  
A;Reference number: 223045; MUID:99051446; PMID:9831526  
A;Accession: T46548  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-257 <ICH>  
A;Cross-references: EMBL:AJ011500; PIDN:CAA09664.1  
A;Experimental source: strain Tu22  
C;Genetics:  
A;Note: gra-orf37

Query Match 1.7%; Score 7; DB 2; Length 257;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 GSGALDA 278  
Db 224 GSGALDA 230  
|||||

# RESULT 63

A75618  
Probable chromosome partitioning ATPase, ParA family - Deinococcus radiodurans (strain R  
C;Species: Deinococcus radiodurans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
A;Accession: A75618  
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896; PMID:10567266  
A;Accession: A75618  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-258 <WHI>  
A;Cross-references: GB:AE001826; NID:G6460827; PIDN:AAF12609.1; PID:G6460905; TIGR:DRB00  
A;Experimental source: strain R1  
C;Genetics:  
A;Gene: DRB0001  
A;Map position: megaplasmid  
A;Genome: plasmid  
A;Note: plasmid MPI  
C;Superfamily: regulatory protein spo0J

Query Match 1.7%; Score 7; DB 2; Length 258;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 341 LGQLAAA 347  
Db 127 LGQLAAA 133  
|||||

# RESULT 64

G83245  
Probable ATP-binding component of ABC transporter PA3212 [imported] - Pseudomonas aerugi  
C;Species: Pseudomonas aeruginosa  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
A;Accession: G83245  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A;Reference number: A82950; MUID:20437337; PMID:10984043  
A;Accession: G83245  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-264 <STO>  
A;Cross-references: GB:AE004744; GB:AE004091; NID:g9949317; PIDN:AAG06600.1; GSPDB:GN001  
A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA3212

Query Match 1.7%; Score 7; DB 2; Length 264;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 138 VAATDDA 144  
Db 234 VAATDDA 240  
|||||

# RESULT 65

T10644  
Hypothetical protein T13K14.180 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 15-Oct-1999  
A;Accession: T10644  
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lenc  
submitted to the Protein Sequence Database, June 1999  
A;Reference number: Z16991  
A;Accession: T10644  
A;Molecule type: DNA  
A;Residues: 1-266 <BEV>  
A;Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:T13K14.180  
A;Experimental source: cultivar Columbia; BAC clone T13K14  
C;Genetics:  
A;Gene: ATSP:T13K14.180  
A;Map position: 4  
A;Introns: 47/3

Query Match 1.7%; Score 7; DB 2; Length 266;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 KAPATLA 97  
Db 22 KAPATLA 28  
|||||

# RESULT 66

S72844  
beta-aspartate methyltransferase pimT - Mycobacterium leprae  
N;Alternate names: B2126 Ci 165 protein  
C;Species: Mycobacterium leprae  
C;Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 23-Mar-2001  
A;Accession: S72844  
R;Smith, D.R.; Robinson, K.  
submitted to the EMBL Data Library, November 1993  
A;Description: Mycobacterium leprae cosmid B2126.  
A;Reference number: S72585  
A;Accession: S72844  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-280 <SMI>  
A;Cross-references: EMBL:U00017; NID:g466994; PIDN:AAA17184.1; PID:g466999  
C;Genetics:  
A;Gene: pimT  
A;Start codon: GTG

Query Match 1.7%; Score 7; DB 2; Length 280;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 270 GAGSGAL 276  
Db 107 GAGSGAL 113  
|||||

# RESULT 67

A70513  
Hypothetical protein RV2118c - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: A70513  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: A70513  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-280 <COL>  
A;Cross-references: GB:Z97559; GB:AL123456; NID:g3261820; PIDN:CAB10715.1; PID:e328642;  
A;Experimental source: strain H37Rv  
C;Genetics:  
A;Gene: RV2118c

Query Match 1.7%; Score 7; DB 2; Length 280;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 270 GAGSGAL 276  
Db 107 GAGSGAL 113

RESULT 68  
A61047  
ectodermal (ect) - fruit fly (*Drosophila melanogaster*) (strain Oregon-R)  
C;Species: *Drosophila melanogaster*  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 16-Feb-1997  
C;Accession: A61047  
R;Raha, D.; Nguyen, Q.D.; Garen, A.  
Dev. Genet. 11, 310-317, 1990  
A;Title: Molecular and developmental analyses of the protein encoded by the *Drosophila* gene *ect*  
A;Reference number: A61047; MUID:91215866; PMID:2090376  
A;Accession: A61047  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-280 <RAH>  
A;Note: In the nucleotide sequence codons for residues 25-28 are not shown and those for residues 29-32 are not shown. Translation is shown for the codon GAG at residue 277 and CTG at residue 278.  
C;Genetics:  
A;Gene: FlyBase:ect  
A;Cross-references: FlyBase:FBgn0000451

Query Match 1.7%; Score 7; DB 2; Length 280;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 QAAPAAAT 137  
Db 95 QAAPAAAT 101

RESULT 69  
C87073  
conserved hypothetical protein ML1313 [imported] - *Mycobacterium leprae*  
C;Species: *Mycobacterium leprae*  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C;Accession: C87073  
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holt, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Mangan, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Massive gene decay in the leprosy bacillus.  
A;Reference number: A86909; MUID:21128732; PMID:11234002  
A;Accession: C87073  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-281 <STO>

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: A70513  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: A70513  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-280 <COL>  
A;Cross-references: GB:Z97559; GB:AL123456; NID:g3261820; PIDN:CAB10715.1; PID:e328642;  
A;Experimental source: strain H37Rv  
C;Genetics:  
A;Gene: RV2118c

Query Match 1.7%; Score 7; DB 2; Length 280;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 270 GAGSGAL 276  
Db 107 GAGSGAL 113

RESULT 70  
A43749  
lignin beta-ether hydrolase (EC 3.3.2.-) ligE - *Pseudomonas paucimobilis*  
C;Species: *Pseudomonas paucimobilis*  
C;Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 08-Oct-1999  
C;Accession: A43749; S35992; S33315  
R;Masai, E.; Katayama, Y.; Kawai, S.; Nishikawa, S.; Yamasaki, M.; Morohoshi, N.  
J. Bacteriol. 173, 7950-7955, 1991  
A;Title: Cloning and sequencing of the gene for a *Pseudomonas paucimobilis* enzyme that degrades lignin  
A;Reference number: A43749; MUID:92078104; PMID:1744051  
A;Accession: A43749  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-281 <MAS1>  
A;Cross-references: EMBL:M69107; NID:g151330; PIDN:AAA25878.1; PID:g151331  
R;Masai, E.; Katayama, Y.; Kubota, S.; Kawai, S.; Yamasaki, M.; Morohoshi, N.  
submitted to DDBJ, September 1992  
A;Description: A bacterial enzyme degrading the model lignin compound beta-etherase is encoded by the *ligE* gene.  
A;Reference number: S35992  
A;Accession: S35992  
A;Molecule type: DNA  
A;Residues: 1-281 <MAS2>  
A;Cross-references: EMBL:D11473; NID:g216887; PIDN:BAA02032.1; PID:d1002514; PID:g21688  
R;Masai, E.; Katayama, Y.; Kubota, S.; Kawai, S.; Yamasaki, M.; Morohoshi, N.  
PDB Lett. 323, 135-140, 1993  
A;Title: A bacterial enzyme degrading the model lignin compound beta-etherase is a merodimer  
A;Reference number: S33313; MUID:93265920; PMID:8495726  
A;Accession: S33315  
A;Molecule type: DNA  
A;Residues: 1-13 <MAS3>  
A;Cross-references: EMBL:D11473  
C;Genetics:  
A;Gene: ligE  
C;Keywords: ether hydrolase

Query Match 1.7%; Score 7; DB 2; Length 281;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 107 ASVARTP 113  
Db 199 ASVARTP 205

RESULT 71  
AG3297  
biopolymer transport exbB protein [imported] - *Brucella melitensis* (strain 16M)  
C;Species: *Brucella melitensis*  
C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 17-May-2002  
C;Accession: AG3297  
R;DelVecchio, V.G.; Kapural, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.; Mazur, M.; Goltzman, E.; Selkov, E.; Eizer, P.H.; Hagius, S.; O'Callaghan, D.; Lete, J.; Mazur, M.; Goltzman, E.; Selkov, E.; Eizer, P.H.; Hagius, S.; O'Callaghan, D.; Lete, J.  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A;Reference number: AD3252; PMID:11756688  
A;Accession: AG3297  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-290 <KUR>  
A;Cross-references: GB:AE008917; PIDN:AAL51546.1; PID:gl7982264; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMEI0365

A:Map position: 1

C:Superfamily: biopolymer transport protein

Query Match 1.7%; Score 7; DB 2; Length 290;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 AAPVAPA 106

|||||

Db 28 AAPVAPA 34

RESULT 72

Tl8584

F15C11.2 - Caenorhabditis elegans (fragment)

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: Tl8584

R:Barlow, K.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z18990

A:Accession: Tl8584

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-292 <WIL>

A:Cross-references: EMBL:Z99266; PIDN:CAB16464.1

A:Experimental source: Clone VF15C11L

C:Genetics:

A:Introns: 31/3; 67/1; 169/3; 212/3

Query Match

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 GKILKDE 56

|||||

Db 54 GKILKDE 60

RESULT 73

C96961

cell division protein (ftsX) [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C:Accession: C96961

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: C96961

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-301 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK78478.1; PID:gl5023359; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC0498

Query Match

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 AATVAAT 141

|||||

Db 28 AATVAAT 34

RESULT 74

F81318

probable galactosyltransferase Cj1139c [imported] - Campylobacter jejuni (strain NCTC

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 06-Jan-2003

C:Accession: F81318

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chill

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barr

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals h

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: F81318

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-303 <PAR>

A:Cross-references: GB:AL139077; GB:AL111168; NID:G6968444; PIDN:CAB73394.1; PID:G6968

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj1139c

C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 ILDMGGG 173

|||||

Db 106 ILDMGGG 112

RESULT 75

F87629

hypothetical protein CC3072 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: F87629

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kol

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: F87629

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-303 <STO>

A:Cross-references: GB:AE005673; NID:gl3424722; PIDN:AAK25034.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC3072

Query Match

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 PVAPAAAS 108

|||||

Db 59 PVAPAAAS 65

Search completed: December 17, 2003, 06:35:56

Job time : 24.3376 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:21:26 ; Search time 11.5265 Seconds  
(without alignments)  
1652.347 Million cell updates/sec

Title: US-09-805-550-2

Perfect score: 405

Sequence: 1 MKNLVKTLKGTNFEIEASPD.....BELTANYLLDHGHEPDDQQQ 405

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.2	240	1 VARN_VIBAN	P74946 vibrio angu
2	9	2.2	264	1 RSPA_STRPU	P46771 strongyloce
3	9	2.2	674	1 PTGA_CORGL	O45298 corynebacte
4	8	2.0	172	1 RL10_RHIL0	Q98n68 rhizobium l
5	8	2.0	254	1 PDHR_ECOLI	P06957 escherichia
6	8	2.0	254	1 PDHR_SALTY	Q914b9 salmonella
7	8	2.0	266	1 STR_KLEPN	P13082 klebsiella
8	8	2.0	289	1 FRAH_ANASP	P46017 anabaena ep
9	8	2.0	346	1 MGB3_HUMAN	O15480 homo sapien
10	8	2.0	393	1 GUN1_USTMA	P54424 ustilago ma
11	8	2.0	464	1 CAP_DICDI	P54654 dictyostell
12	8	2.0	804	1 BCB3_ACEXY	Q9w462 acetobacter
13	8	2.0	1235	1 DPOL_HSV11	P04293 herpes simp
14	7	1.7	124	1 SGP2_CHRV1	O52179 chromatium
15	7	1.7	126	1 CU24_ARADI	P80516 araneus dia
16	7	1.7	154	1 RIB2_LACLA	Q9cgu6 lactococcu
17	7	1.7	156	1 RIB2_PHOLE	Q93a92 photobacter
18	7	1.7	183	1 RRP_TREPA	O83613 treponema p
19	7	1.7	211	1 KGUA_STRP3	Q8p001 streptococc
20	7	1.7	211	1 KGUA_STRP3	Q8p001 streptococc
21	7	1.7	252	1 OLPI_LYCES	Q99ym5 streptococc
22	7	1.7	256	1 HIS4_SYNY3	O41350 lycopersico
23	7	1.7	256	1 TNR9_MOUSE	P74561 synechocyst
24	7	1.7	280	1 LIGE_PSEPA	P20334 mus musculu
25	7	1.7	314	1 NARA_HUMAN	P27457 pseudomonas
26	7	1.7	314	1 NARA_HUMAN	Q93070 homo sapien
27	7	1.7	324	1 NARA_PANTR	Q95ne0 pan troglod
28	7	1.7	324	1 CAHC_HORVU	P40880 hordeum vul
29	7	1.7	335	1 G3P_HALN1	Q9hsr7 halobacteri
30	7	1.7	358	1 G80_GEOCY	Q9x3v3 geodia cydo
31	7	1.7	360	1 POSI_BOVIN	Q97552 bos taurus
32	7	1.7	364	1 INC2_ECOLI	P07673 escherichia
33	7	1.7	365	1 LA01_PONPY	P16211 pongo pygma
33	7	1.7	366	1 T2A1_ACICA	P24546 acinetobact

34	7	1.7	367	1 SGK2_MOUSE	Q9qz85 mus musculu
35	7	1.7	378	1 ADHP_RABIT	O46649 oryctolagus
36	7	1.7	386	1 SUCC_CHLTR	O84928 chlamydia t
37	7	1.7	395	1 EFTU_XYLPA	Q9p9g9 xylaria fas
38	7	1.7	400	1 FXQ1_MOUSE	O70220 mus musculu
39	7	1.7	400	1 PROV_ECOLI	P14755 escherichia
40	7	1.7	400	1 PROV_SALTY	P17328 salmonella
41	7	1.7	427	1 SGK2_HUMAN	Q9hby8 homo sapien
42	7	1.7	448	1 YAP1_CHICK	P46936 gallus gall
43	7	1.7	465	1 ANT3_MOUSE	P32261 mus musculu
44	7	1.7	465	1 FXD3_MOUSE	O61060 mus musculu
45	7	1.7	484	1 COLB_MOUSE	Q9wm33 mus musculu
46	7	1.7	486	1 GALT_LACCA	O84904 lactobacilli
47	7	1.7	498	1 YM22_CAEBL	Q21339 caenorhabdi
48	7	1.7	500	1 DHA1_MOUSE	P25459 mus musculu
49	7	1.7	500	1 DHA1_RAT	P51647 rattus norv
50	7	1.7	500	1 DHAC_RAT	P13601 rattus norv
51	7	1.7	507	1 IRX3_MOUSE	P81067 mus musculu
52	7	1.7	513	1 NQOD_PARDE	P29925 paracoccus
53	7	1.7	524	1 THS3_HALVO	Q9hha2 halobacteri
54	7	1.7	527	1 PTB_MOUSE	P17225 mus musculu
55	7	1.7	531	1 PTB_HUMAN	P26599 homo sapien
56	7	1.7	537	1 GAG_BAIVM	P03341 baboon endo
57	7	1.7	548	1 SYK_RHILO	Q98bc8 rhizobium l
58	7	1.7	551	1 SYK_BRUME	Q8yem8 brucella me
59	7	1.7	551	1 SYK_BRUSU	P59225 brucella su
60	7	1.7	555	1 PTB_RAT	Q00438 rattus norv
61	7	1.7	557	1 PTB_PIG	Q29099 sus scrofa
62	7	1.7	559	1 YHUJ_ECOLI	P37659 escherichia
63	7	1.7	597	1 Y745_ARATH	P59278 arabidopsis
64	7	1.7	629	1 SYR_PYRHO	O59147 pyrococcus
65	7	1.7	630	1 COG6_DROME	Q9v564 drosophila
66	7	1.7	641	1 PRIM_UREPA	Q9p266 ureaplasma
67	7	1.7	646	1 VP40_HSVB	P28336 equine herp
68	7	1.7	660	1 PKM1_COREF	Q8fuis corynebacte
69	7	1.7	666	1 SCRT_DROME	P45843 drosophila
70	7	1.7	682	1 TUP1_KLULA	P56094 kluyveromyc
71	7	1.7	694	1 EFG_BRUME	Q8vhp3 brucella me
72	7	1.7	738	1 STII_YEAST	P23561 saccharomyc
73	7	1.7	750	1 ELS_CHICK	P07916 gallus gall
74	7	1.7	775	1 LYS4_EMENI	Q92412 emericella
75	7	1.7	777	1 FRZE_MYXPO	P18769 myxococcus
76	7	1.7	801	1 VP34_SCHPO	P50520 schizosacch
77	7	1.7	812	1 P3K2_SOYBN	P42348 glycine max
78	7	1.7	814	1 P3K1_SOYBN	P42347 glycine max
79	7	1.7	814	1 P13K_ARATH	P42339 arabidopsis
80	7	1.7	816	1 P3K4_DICDI	P54676 dictyostell
81	7	1.7	857	1 APPI_SCHPO	Q9p7e8 schizosacch
82	7	1.7	870	1 BCAL_HUMAN	P56945 homo sapien
83	7	1.7	874	1 BCAL_MOUSE	Q61140 mus musculu
84	7	1.7	886	1 SYA_BARBA	P70865 bartonella
85	7	1.7	888	1 SYA_ZYMO	Q9rnn8 zymomonas m
86	7	1.7	917	1 AKH2_MAIZE	P49080 zea mays (m
87	7	1.7	920	1 AKH1_MAIZE	P49079 zea mays (m
88	7	1.7	964	1 MSH2_YEAST	P25847 saccharomyc
89	7	1.7	968	1 BCAL_RAT	Q63767 rattus norv
90	7	1.7	980	1 UBE4_CAEBL	Q09349 caenorhabdi
91	7	1.7	1004	1 PHC1_HUMAN	P78364 homo sapien
92	7	1.7	1012	1 PHC1_MOUSE	Q64028 mus musculu
93	7	1.7	1020	1 VP34_CANAL	Q92213 candida alb
94	7	1.7	1047	1 ANPB_BOVIN	P46197 bos taurus
95	7	1.7	1047	1 ANPB_HUMAN	P20594 homo sapien
96	7	1.7	1047	1 ANPB_RAT	P16067 rattus norv
97	7	1.7	1049	1 DSG1_HUMAN	Q02413 homo sapien
98	7	1.7	1055	1 POL_HV2ST	P20876 human immun
99	7	1.7	1061	1 RNE_ECOLI	P21513 escherichia
100	7	1.7	1070	1 Y355_HUMAN	O15063 homo sapien

ALIGNMENTS

RESULT 1

```

VANR VIBAN
ID VANR VIBAN STANDARD; PRT; 240 AA.
AC P74946;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcriptional activator protein vanR.
GN VANR
OS Vibrio anguillarum (Listonella anguillarum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Listonella.
OX NCBI_TaxID=55601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NB10 / Serotype O1;
RA MEDLINE=97284506; PubMed=9139920;
RA Milton D.L., Hardman A., Camara M., Chhabra S.R., Bycroft B.W.,
RA Stewart G.S.A.B., Williams P.;
RT "Quorum sensing in Vibrio anguillarum: characterization of the
RT vanI/vanR locus and identification of the autoinducer N-(3-
RT oxodecanoyl)-L-homoserine lactone.";
RL J. Bacteriol. 179:3004-3012(1997).
CC -!- FUNCTION: PROBABLE TRANSCRIPTIONAL ACTIVATOR. BINDS TO AUTOINDUCER
CC MOLECULE OHL.
CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS. AUTOINDUCER REGULATED SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; U69677; AAC45213.1; -.
DR InterPro; IPR005143; Autoind bind.
DR InterPro; IPR000792; HTH LuxR.
DR Pfam; PF03472; Autoind_bind; 1.
DR Pfam; PF00196; GerE; 1.
DR PRINTS; PR00038; HTHLUXR.
DR ProDom; PD000307; HTH LuxR; 1.
DR SMART; SM00421; HTH_LUXR; 1.
DR PROSITE; PS00622; HTH_LUXR_FAMILY; FALSE NEG.
KW Transcription regulation; DNA-binding; Activator; Quorum sensing.
FT DNA BIND 193 212 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 240 AA; 27253 MW; BF9E0C40564C6663 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.66; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 313 ILRLIQENQ 321
DB 4 ILRLIQENQ 12

RESULT 2
ID RSP4 STRPU STANDARD; PRT; 264 AA.
AC P46771;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 40S ribosomal protein SA (P40) (34/67 kDa laminin binding protein)
DE (Fragment).
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.

VANR VIBAN
ID VANR VIBAN STANDARD; PRT; 240 AA.
AC P74946;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcriptional activator protein vanR.
GN VANR
OS Vibrio anguillarum (Listonella anguillarum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Listonella.
OX NCBI_TaxID=55601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NB10 / Serotype O1;
RA MEDLINE=97284506; PubMed=9139920;
RA Milton D.L., Hardman A., Camara M., Chhabra S.R., Bycroft B.W.,
RA Stewart G.S.A.B., Williams P.;
RT "Quorum sensing in Vibrio anguillarum: characterization of the
RT vanI/vanR locus and identification of the autoinducer N-(3-
RT oxodecanoyl)-L-homoserine lactone.";
RL J. Bacteriol. 179:3004-3012(1997).
CC -!- FUNCTION: PROBABLE TRANSCRIPTIONAL ACTIVATOR. BINDS TO AUTOINDUCER
CC MOLECULE OHL.
CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS. AUTOINDUCER REGULATED SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; U69677; AAC45213.1; -.
DR InterPro; IPR005143; Autoind bind.
DR InterPro; IPR000792; HTH LuxR.
DR Pfam; PF03472; Autoind_bind; 1.
DR Pfam; PF00196; GerE; 1.
DR PRINTS; PR00038; HTHLUXR.
DR ProDom; PD000307; HTH LuxR; 1.
DR SMART; SM00421; HTH_LUXR; 1.
DR PROSITE; PS00622; HTH_LUXR_FAMILY; FALSE NEG.
KW Transcription regulation; DNA-binding; Activator; Quorum sensing.
FT DNA BIND 193 212 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 240 AA; 27253 MW; BF9E0C40564C6663 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.66; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 313 ILRLIQENQ 321
DB 4 ILRLIQENQ 12

RESULT 2
ID RSP4 STRPU STANDARD; PRT; 264 AA.
AC P46771;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 40S ribosomal protein SA (P40) (34/67 kDa laminin binding protein)
DE (Fragment).
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.

MEDLINE=96042034; PubMed=7589249;
RA Hung M., Rosenthal E.T., Boblett B., Benson S.;
RT "Characterization and localized expression of the laminin binding
RT protein/p40 (LBP/p40) gene during sea urchin development.";
Exp. Cell Res. 221:221-230(1995).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE S2P FAMILY OF RIBOSOMAL PROTEINS.
CC -!- CAUTION: WAS ORIGINALLY THOUGHT TO BE A LAMININ RECEPTOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; U02369; AAA90976.1; -.
DR InterPro; IPR001865; Ribosomal_S2.
DR Pfam; PF00318; Ribosomal_S2; 1.
DR PROSITE; PS00962; RIBOSOMAL_S2_1; PARTIAL.
DR PROSITE; PS00963; RIBOSOMAL_S2_2; 1.
KW Ribosomal protein.
FT NON TER 1 1
SQ SEQUENCE 264 AA; 28141 MW; D0AB883C051CF859 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 264;
Best Local Similarity 100.0%; Pred. No. 0.72; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 99 PAAPVAPAA 107
DB 227 PAAPVAPAA 235

RESULT 3
PTGA CORGL
ID PTGA CORGL STANDARD; PRT; 674 AA.
AC Q45298;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PTS system, glucose-specific IIAAC component (EIIABC-GLC) (Glucose-
DE permease IIAAC component) (Phosphotransferase enzyme II, ABC
DE component) (EC 2.7.1.69) (EII-GLC/EIIC-GLC).
GN PTSG.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OX Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OX Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13869;
RA Yoon K.-H.;
RT "Cloning and nucleotide sequence of enzyme II of Brevibacterium
RT lactofermentum phosphotransferase system.";
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOTRANSFERASE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Contains 1 PTS EIIA domain.
CC -!- SIMILARITY: Contains 1 PTS EIIB domain.
CC -!- SIMILARITY: Contains 1 PTS EIIC domain.
CC -!- SIMILARITY: Contains 1 PTS EIIC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L18875; AAA22992.1; -;  
 DR HSP; P08837; IGIC.  
 DR InterPro; IPR001127; PTS\_EI1A.  
 DR InterPro; IPR001996; PTS\_EI1B.  
 DR InterPro; IPR003352; PTS\_EI1C.  
 DR Pfam; PF00358; PTS\_EI1A\_1; 1.  
 DR Pfam; PF00367; PTS\_EI1B; 1.  
 DR Pfam; PF02378; PTS\_EI1C; 1.  
 DR ProDom; PD002243; PTS\_EI1A; 1.  
 DR ProDom; PD001476; PTS\_EI1B; 1.  
 DR TIGRFAMs; TIGR00830; PTS\_EI1A\_1; 1.  
 DR PROSITE; PS00371; PTS\_EI1A\_1; 1.  
 DR PROSITE; PS01035; PTS\_EI1B\_CYS; 1.  
 KW Phosphotransferase system; Sugar transport; Transferase;  
 KW Phosphorylation; Transmembrane; Plasmid.  
 FT DOMAIN 1 43  
 FT DOMAIN ?  
 FT DOMAIN 542 674  
 FT TRANSMEM 126 146  
 FT TRANSMEM 162 182  
 FT TRANSMEM 193 213  
 FT TRANSMEM 225 245  
 FT TRANSMEM 260 280  
 FT TRANSMEM 303 323  
 FT TRANSMEM 344 364  
 FT TRANSMEM 377 397  
 FT TRANSMEM 409 429  
 FT TRANSMEM 442 462  
 FT TRANSMEM 493 513  
 FT MOD RES 28 28  
 FT MOD RES 594 594  
 SQ SEQUENCE 674 AA; 71626 MW; 67A75AAF76E42FA2 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 674;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 PAAPVAPAA 107  
 |||||  
 Db 490 PAAPVAPAA 498

RESULT 4  
 RL10\_RH10  
 ID\_RH10 RH10 STANDARD; PRT; 172 AA.  
 AC Q98N68;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE 50S ribosomal protein L10.  
 GN RPLJ OR MR0274  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11211968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti.";

RL DNA Res. 7:331-338(2000).  
 CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AP002994; BAB47894.1; -;  
 DR HAWAP; MF\_00362; -; 1.  
 DR InterPro; IPR001790; Ribosomal L10.  
 DR InterPro; IPR002363; Ribosomal L10eub.  
 DR Pfam; PF00466; Ribosomal L10; 1.  
 DR PROSITE; PS01109; RIBOSOMAL\_L10; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 172 AA; 17734 MW; 99AE3D33556AA5ED CRC64;

Query Match 2.0%; Score 8; DB 1; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 4.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 APAASVAR 111  
 |||||  
 Db 153 APAASVAR 160

RESULT 5  
 PDHR\_ECOLI  
 ID\_PDHR\_ECOLI STANDARD; PRT; 254 AA.  
 AC P06957; O53381;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Pyruvate dehydrogenase complex repressor.  
 GN PDHR OR GENA OR ACEC OR B0113 OR C0140 OR Z0123 OR ECS0117.  
 OS Escherichia coli.  
 OS Escherichia coli O6, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562, 217992, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=83209630; PubMed=6343085;  
 RA Stephens P.E., Darlison M.G., Lewis H.M., Guest J.R.;  
 RA "The pyruvate dehydrogenase complex of Escherichia coli K12.  
 RT Nucleotide sequence encoding the pyruvate dehydrogenase component.";  
 RL Eur. J. Biochem. 133:155-162(1983).  
 RN [2]  
 RP REVISIONS.  
 RA Quail M.A.;  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glaeser J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RA "The complete genome sequence of Escherichia coli K-12.";  
 RT Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Weich R.A., Burland V., Plunkett G. III, Redford P., Roessch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,



RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic *Escherichia coli*";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 RN [5]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=O157:H7 / BDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,  
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";  
 RL Nature 409:529-533 (2001).  
 RN [6]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RA "Complete genome sequence of enterohemorrhagic *Escherichia coli*  
 O157:H7 and genomic comparison with a laboratory strain K-12";  
 RT DNA Res. 8:11-22 (2001).  
 RL [7]  
 RN CHARACTERIZATION, AND VARIANT ACEC816.  
 RP STRAIN=K12;  
 RC MEDLINE=94085588; PubMed=8262214;  
 RX Haydon D.J., Quail M.A., Guest J.R.;  
 RA "A mutation causing constitutive synthesis of the pyruvate  
 dehydrogenase complex in *Escherichia coli* is located within the *pdhR*  
 gene";  
 RT FEBS Lett. 336:43-47 (1993).  
 RL  
 CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE PYRUVATE DEHYDROGENASE  
 CC COMPLEX GENES ACEEF AND LPD  
 CC -!- SIMILARITY: BELONGS TO THE GNTR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; V01498; CAA24739.1; -;  
 DR EMBL; D26562; BAB96682.1; ALT\_SEQ.  
 DR EMBL; A5000120; AAC73224.1; -;  
 DR EMBL; A5016755; AAN78636.1; ALT\_INIT.  
 DR EMBL; A5005187; AAG54417.1; -;  
 DR EMBL; AP002550; BAB33540.1; -;  
 DR EMBL; S67363; AAB29356.1; -;  
 DR PIR; A64734; BVECA.  
 DR PIR; E85494; E85494.  
 DR PIR; E90643; E90643.  
 DR EcoGene; EG11088; *pdhR*.  
 DR InterPro; IPR000524; HTH\_Gntr.  
 DR Pfam; PF00392; *gntr*; 1.  
 DR PRINTS; PR00035; HTHGNTR.  
 DR SMART; SM00345; HTH GNTR; 1.  
 DR PROSITE; PS00043; HTH GNTR FAMILY; 1.  
 KW Transcription regulation; DNA-binding; Repressor; Complete proteome.  
 FT DNA\_BIND 37 56 H-T-H MOTIF (BY SIMILARITY).  
 FT CONFLICT 118 118 R -> C (IN ACEC816; CONSTITUTIVELY  
 ACTIVATED).  
 FT  
 SQ SEQUENCE 254 AA; 29425 MM; AF20245D33F751D CRC64;  
 Query Match 2.0%; Score 8; DB 1; Length 254;  
 Best Local Similarity 100.0%; Pred. No. 6.5;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 359 REAIQRLQ 366  
 Db 52 REAIQRLQ 59  
 RESULT 6  
 PDHR\_SALTY STANDARD; PRT; 254 AA.  
 ID PDHR\_SALTY STANDARD; PRT; 254 AA.  
 AC Q9L4H9;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Pyruvate dehydrogenase complex repressor.  
 GN PDHR OR STM0151 OR STY0174 OR TO157.  
 OS *Salmonella typhimurium*, and  
 OS *Salmonella typhi*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; *Salmonella*.  
 OX NCBI\_TaxID=602, 601;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=S.typhimurium; STRAIN=SL1344;  
 RA Cano D., Casadesus J., Garcia-del Portillo F.;  
 RT "Characterization of a *Salmonella*-specific region located between *ampE*  
 and *aroP* genes";  
 RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 Waterston R., Wilson R.K.;  
 RA "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 LT2";  
 RT Nature 413:852-856 (2001).  
 RL [3]  
 RN SEQUENCE FROM N.A.  
 RP SPECIES=S.typhi; STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 Churcher C., Mungall K.B., Bentley S.D., Holden M.T.G., Sebahia M.,  
 Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 Whitehead S., Barrall B.G.;  
 RA "Complete genome sequence of a multiple drug resistant *Salmonella*  
 enterica serovar Typhi CT18";  
 RT Nature 413:848-852 (2001).  
 RL [4]  
 RN SEQUENCE FROM N.A.  
 RP SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;  
 RX MEDLINE=22531367; PubMed=12644504;  
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
 Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
 RA "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2  
 and CT18";  
 RT J. Bacteriol. 185:2330-2337 (2003).  
 RL CC  
 CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE PYRUVATE DEHYDROGENASE  
 CC COMPLEX GENES ACEEF AND LPD (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE GNTR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AJ242516; CAB98840.1; -;  
 DR EMBL; AE008701; AAL19115.1; -;  
 DR EMBL; AL627265; CAD01310.1; -;  
 DR EMBL; AE016834; AAO67889.1; -;  
 DR StyGene; SG77777; pdhr.  
 DR InterPro; IPR000524; HTH\_Gntr.  
 DR Pfam; PF00392; gntR; 1.  
 DR PRINTS; PR00035; HTHGNTR.  
 DR SMART; SM00345; HTH\_GNTR; 1.  
 DR PROSITE; PS00043; HTH\_GNTR\_FAMILY; 1.  
 KW Transcription regulation; DNA-binding; Repressor; Complete proteome.  
 FT DNA\_BIND 37 56 H-T-H MOTIF (BY SIMILARITY).  
 FT CONFLICT 135 135 E -> K (IN REF. 1).  
 SQ SEQUENCE 254 AA; 29411 MW; 990A66833052EAB3 CRC64;

Query Match 2.0%; Score 8; DB 1; Length 254;  
 Best Local Similarity 100.0%; Pred. No. 6.5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRL 366  
 |||||  
 Db 52 REAIQRL 59

## RESULT 7

ID STR\_KLEPN STANDARD; PRT; 266 AA.  
 AC P13082;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE Streptomycin 3',-kinase (EC 2.7.1.87) (Streptomycin 3',-  
 DE phosphotransferase) (SPH).  
 GN STR.  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Klebsiella.  
 OX NCBI\_TaxID=573;  
 RN [1]

SEQUENCE FROM N.A.  
 RX MEDLINE=85215465; PubMed=3889831;  
 RA Mazodier P., Cossart P., Giraud E., Gasser F.;  
 RT "Completion of the nucleotide sequence of the central region of Tn5  
 RT confirms the presence of three resistance genes.";  
 RL Nucleic Acids Res. 13:195-205(1985).  
 CC -!- FUNCTION: THE AMINOGLYCOSIDE PHOSPHOTRANSFERASES ACHIEVE  
 CC INACTIVATION OF THEIR ANTIBIOTIC SUBSTRATES BY PHOSPHORYLATION.  
 CC -!- CATALYTIC ACTIVITY: ATP + streptomycin = ADP + streptomycin 3',-  
 CC phosphate.  
 CC -!- MISCELLANEOUS: THIS ENZYME IS ENCODED BY THE KANAMYCIN AND  
 CC NEOMYCIN RESISTANCE TRANSPOSON TNS.  
 CC -!- SIMILARITY: TO OTHER AMINOGLYCOSIDE PHOSPHOTRANSFERASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U00004; AAA73392.1; -;  
 DR EMBL; X01702; CAA25854.1; -;  
 DR InterPro; IPR006748; APH\_6\_hur.  
 DR Pfam; PF04655; APH\_6\_hur; 1.  
 KW Antibiotic resistance; Transferase; Kinase; ATP-binding;  
 KW Transposable element.  
 FT ACT\_SITE 154 154 BY SIMILARITY.  
 SQ SEQUENCE 266 AA; 29066 MW; 30A9DA5D0E1A826C CRC64;

Query Match 2.0%; Score 8; DB 1; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 6.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 APAASVAR 111  
 |||||  
 Db 132 APAASVAR 139

## RESULT 8

ID FRAH\_ANASP STANDARD; PRT; 289 AA.  
 AC P46017;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE FRAH protein.  
 GN FRAH OR ALR1603.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bauer C.C., Scappino L., Haselkorn R.;  
 RT "Bacterial subtracted cDNA libraries containing genes involved in the  
 RT differentiation of vegetative cells to heterocysts allow a new twist  
 RT on an old method of isolating developmental genes.";  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]

SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 CC -!- FUNCTION: PUTATIVE HETEROCYST TO VEGETATIVE CELL CONNECTION.  
 CC -!- SIMILARITY: Contains 1 FHA domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U14553; AAA50356.1; -;  
 DR EMBL; AP003586; BAB77969.1; -;  
 DR PIR; AE2006; AE2006.  
 DR InterPro; IPR000253; FHA.  
 DR Pfam; PF00498; FHA; 1.  
 DR SMART; SM00240; FHA; 1.  
 DR PROSITE; PS50006; FHA\_DOMAIN; 1.  
 KW Heterocyst; Zinc-finger; Complete proteome.  
 FT ZN\_FING 18 48 POTENTIAL.  
 FT DOMAIN 67 172 PRO-RICH.  
 FT DOMAIN 204 260 FHA.  
 SQ SEQUENCE 289 AA; 29922 MW; 47C9091E50B39716 CRC64;

Query Match 2.0%; Score 8; DB 1; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 7.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 AAPAATVA 139  
 |||||  
 Db 56 AAPAATVA 63

## RESULT 9

MGB3 HUMAN  
 ID MGB3\_HUMAN STANDARD; PRT; 346 AA.  
 AC O15480; O75861;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Melanoma-associated antigen B3 (MAGE-B3 antigen).  
 GN MAGEB3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98110575; PubMed=9441743;  
 RA Lurquin C., de Smet C., Brasseur F., Muscatelli F., Martelange V.,  
 RA de Plaen E., Brasseur R., Monaco A.P., Boon T.;  
 RT "Two members of the human MAGEB gene family located in Xp21.3 are  
 RT expressed in tumors of various histological origins";  
 RL Genomics 46:397-408(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Muzny D., Arenson A.D., Adams C., Brundage E., Bunac C., Carvelli K.,  
 RA Chacko J., Chen J., Di W., Ding Y., Dugan S., Durbin J., Forcum J.,  
 RA Ganesh R., Garcia C., Goodman M., Gorrell J.H., Haywood M.,  
 RA Hernandez J., Jackson L., Jin S., Kampal R., Karpathy S., Kovar C.,  
 RA Leal B., Li Y., Lichtarge O., Liu W., Logan O., Lu J., Ly T.,  
 RA Martinez C., Oswal G., Perez L., Rashid N.D., Rowland K., Savage L.,  
 RA Scherer S.E., Shen H., Simon K., Stovall K., Timms K.M., Todd J.,  
 RA Vo O., Williamson A., Worley K.C., Yu W., Chinault C., Nelson D.,  
 RA Gibbs R.A.;  
 RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RL  
 CC -1- TISSUE SPECIFICITY: Expressed in testis.  
 CC -1- SIMILARITY: Contains 1 MAGE domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U93163; AAC23618.1; --  
 CC DR EMBL; AC005185; AAD10636.1; --  
 CC DR Genew; HGNC:6810; MAGEB3.  
 CC DR MIM; 300152; --  
 CC DR GO; GO:0008222; F: tumor antigen; TAS.  
 CC DR InterPro; IPR002190; MAGE.  
 CC DR Pfam; PF01454; MAGE; 1.  
 CC DR PROSITE; PS00838; MAGE; 1.  
 CC KW Antigen; Multigene family.  
 CC FT DOMAIN 111 310 MAGE.  
 CC FT CONFLICT 107 107 H -> R (IN REF. 2).  
 CC FT CONFLICT 112 112 T -> I (IN REF. 2).  
 CC SQ SEQUENCE 346 AA; 39179 MW; 4F9EABF6D813BA35 CRC64;  
 Query Match 2.0%; Score 8; DB 1; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 75 SKAKSSS 82  
 Db 336 SKAKSSS 343  
 RESULT 10  
 GUN1\_USTMA STANDARD; PRT; 393 AA.  
 AC P54424;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 1)  
 DE (Cellulase 1) (EG 1).  
 GN EGLI.  
 OS Ustilago maydis (Smut fungus).  
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;  
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.  
 OX NCBI\_TaxID=5270;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96145728; PubMed=8590631;  
 RA Schaevecker F., Wanner G., Kahmann R.;  
 RT "Filament-specific expression of a cellulase gene in the dimorphic  
 RT fungus Ustilago maydis";  
 RL Biol. Chem. Hoppe-Seyler 376:617-625(1995).  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: HYPHAL TIP.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN FILAMENTOUS DIKARYON.  
 CC -1- PTM: MAY ALSO BE O-GLYCOSYLATED.  
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL  
 CC HYDROLASES).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; S81598; AAB36147.1; --  
 CC DR FIR; S59499; S59499.  
 CC DR HSP; P43316; 2ENG.  
 CC DR InterPro; IPR000334; Glyco\_hydro 45.  
 CC DR Pfam; PF02015; Glyco\_hydro 45; 1.  
 CC DR PROSITE; PS01140; GLYCOSYL\_HYDROL\_F45; 1.  
 CC KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Glycoprotein.  
 CC FT SIGNAL 1 26 POTENTIAL.  
 CC FT CHAIN 27 393 ENDOGLUCANASE 1.  
 CC FT ACT\_SITE 34 34 NUCLEOPHILE (BY SIMILARITY).  
 CC FT ACT\_SITE 152 152 PROTON DONOR (BY SIMILARITY).  
 CC FT DOMAIN 270 385 ALA/GLY/SER-RICH.  
 CC FT DOMBOHD 343 343 N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC SQ SEQUENCE 393 AA; 39594 MW; 65C753C610CD6AD3 CRC64;  
 Query Match 2.0%; Score 8; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 9.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 259 PSGGSNPG 266  
 Db 362 PSGGSNPG 369  
 RESULT 11  
 CAP\_DICDI STANDARD; PRT; 464 AA.  
 AC P54554;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Adenylyl cyclase-associated protein (CAP).  
 GN CAP.  
 OS Dictyostellium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX3;  
 RX MEDLINE=96228685; PubMed=8688557;

RA Gottwald U., Brokamp R., Karakesisoglou I., Schleicher M.,  
RA Nogel A.A.;  
RT Identification of a cyclase-associated protein (CAP) homologue in  
RT Dictyostelium discoideum and characterization of its interaction with  
RT actin.";  
RL Mol. Biol. Cell 7:261-272(1996).  
CC -!- FUNCTION: MAY HAVE A REGULATORY BIFUNCTIONAL ROLE. BINDS G-ACTIN  
CC AND PIP2. INVOLVED IN MICROFILAMENT REORGANIZATION NEAR THE PLASMA  
CC MEMBRANE IN A PIP2-REGULATED MANNER.  
CC -!- SUBCELLULAR LOCATION: CELL MEMBRANE.  
CC -!- DOMAIN: THE C-TERMINUS IS RESPONSIBLE FOR SEQUESTERING G-ACTIN.  
CC THE N-TERMINUS IS REQUIRED FOR THE PIP2 MODULATION OF CAP  
CC FUNCTION.  
CC -!- SIMILARITY: BELONGS TO THE CAP FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U43027; AAB09713.1; -;  
CC DictyDb; DD00012; -;  
CC InterPro; IPR001837; CAP.  
CC InterPro; IPR006599; CAP.  
CC Pfam; PF01213; CAP; 1.  
CC SMART; SMO0673; CAP; 2.  
CC PROSITE; PS01088; CAP\_1; 1.  
CC PROSITE; PS01089; CAP\_2; 1.  
CC Membrane; Actin-binding.  
CC KW DOMAIN 33 42 POLY-SER.  
CC FT DOMAIN 217 253 PRO-RICH.  
CC SQ SEQUENCE 464 AA; 49641 MW; 7ABDD05D8A0148B4 CRC64;  
  
Query Match 2.0%; Score 8; DB 1; Length 464;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 99 PAAPVAPA 106  
DB 234 PAAPVAPA 241  
  
RESULT 12  
BCB3 ACXY  
ID BCB3 ACXY STANDARD; PRT; 804 AA.  
AC Q9XK62;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cyclic di-GMP binding protein precursor (Cellulose synthase regulatory  
DE subunit) (Cellulose synthase protein B) (CDGPB).  
GN BCB3.  
OS Acetobacter xylinus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
OC Acetobacteraceae; Gluconacetobacter.  
OX NCBI\_TaxID=28448;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 7664 / IFO 13693;  
RX MEDLINE=99310341; PubMed=10382968;  
RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,  
RA Inoue Y.;  
RT Cloning of cellulose synthase genes from Acetobacter xylinum JCM  
RT 7664: implication of a novel set of cellulose synthase genes.";  
RL DNA Res. 6:109-115(1999).  
CC -!- FUNCTION: Binds the cellulose synthase activator, bis-(3'-5')  
CC cyclic diguanylic acid (c-di-GMP) (By similarity).  
CC -!- PATHWAY: Bacterial cellulose biosynthesis.  
CC -!- SUBUNIT: Tightly associated with the cellulose synthase catalytic  
CC subunit (By similarity).

CC -!- SIMILARITY: BELONGS TO THE ACSB/BCSB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB015802; BAA77586.1; -;  
CC InterPro; IPR003920; Celleynth\_B.  
CC Pfam; PF031170; BCSB; 1.  
CC PRINTS; PR01440; CELLSYNTHASEB.  
CC Cellulose biosynthesis; Signal.  
CC FT SIGNAL 1 18 POTENTIAL.  
CC FT CHAIN 19 804 CYCLIC DI-GMP BINDING PROTEIN.  
CC SQ SEQUENCE 804 AA; 85510 MW; 1888ADD82EFC50A3 CRC64;  
  
Query Match 2.0%; Score 8; DB 1; Length 804;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 100 AAPVAPAA 107  
DB 41 AAPVAPAA 48  
  
RESULT 13  
DPOL HSV1  
ID DPOL HSV1 STANDARD; PRT; 1235 AA.  
AC P04293;  
DT 20-MAR-1987 (Rel. 04, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DNA polymerase (EC 2.7.7.7).  
GN UL30.  
OS Herpes simplex virus (type 1 / strain 17).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirus.  
OX NCBI\_TaxID=10299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88274327; PubMed=2839594;  
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,  
RA McNab D., Perry L.J., Scott J.E., Taylor P.;  
RT "The complete DNA sequence of the long unique region in the genome of  
RT herpes simplex virus type 1.";  
RL J. Gen. Virol. 69:1531-1574(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86067223; PubMed=2999714;  
RA Quinn J.P., McGeoch D.J.;  
RT "DNA sequence of the region in the genome of herpes simplex virus  
RT type 1 containing the genes for DNA polymerase and the major DNA  
RT binding protein.";  
RL Nucleic Acids Res. 13:8143-8163(1985).  
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + [DNA] (N).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X14112; CAA32323.1; -;  
CC DR EMBL; X03181; CAA26941.1; -;  
CC DR PIR; A00715; DJBEV1.

```
DR PIR; C30085; DJBEH7.
DR InterPro; IPR0061172; DNA_pol_B.
DR InterPro; IPR0061134; DNA_pol_B_dom.
DR InterPro; IPR0061133; DNA_pol_B_exo.
DR Pfam; PF001136; DNA_pol_B; 1.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR SMART; SM00486; POLBc; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
FT DNA-binding; Nuclear protein.
FT CONFLICT 114 126 VLRVGGGFWRPR -> APPRRVGRLLAA (IN
SQ SEQUENCE 1235 AA; 136420 MW; E8CD41D6EDED8343 CRC64;
REF. 2).

Query Match 2.0%; Score 8; DB 1; Length 1235;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 TLAQPAAP 102
DB 325 TLAQPAAP 332
|||||

RESULT 14
SGP2 CHRVI STANDARD; PRT; 124 AA.
AC OS2179;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Sulfur globule protein CV2 precursor.
GS SGPB OR SGP2.
ON Chromatium vinosum.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales;
OC Chromatiaceae; Allochromatium.
OX NCBI_TaxID=1049;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D / ATCC 17899 / DSM 180;
RX MEDLINE=98228197; PubMed=9560425;
RA Pattaragulwanit K., Brune D.C., Trueper H.G., Dahl C.;
RT "Molecular genetic evidence for extracytoplasmic localization of
sulfur globules in Chromatium vinosum.";
RL Arch. Microbiol. 169:434-444(1998).
RN [2]
RP SEQUENCE OF 23-72, AND CHARACTERIZATION.
RC STRAIN=D / ATCC 17899 / DSM 180;
RX MEDLINE=96009136; PubMed=7575095;
RA Brune D.C.;
RT "Isolation and characterization of sulfur globule proteins from
Chromatium vinosum and Thiocapsa roseopersicina.";
RL Arch. Microbiol. 163:191-199(1995).
CC -1- FUNCTION: STRUCTURAL PROTEIN OF THE SULFUR GLOBULES, WHICH ARE
INTRACELLULAR GLOBULES THAT SERVE FOR SULFUR STORAGE IN PURPLE
SULFUR BACTERIA.
CC -1- SUBUNIT: THE PROTEIN ENVELOPE OF THE SULFUR GLOBULES ARE COMPOSED
OF THREE DIFFERENT PROTEINS; CV1, CV2 AND CV3.
CC -1- MASS SPECTROMETRY: MW=10651; METHOD=MALDI; RANGE=22-124.
CC -1- SIMILARITY: TO C.VINOSUM CV1 AND T.ROSEOPERSICINA TR0.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF025952; AAB91547.1;
KW Structural protein; Signal.
FT SIGNAL 1 22
FT CHAIN 23 124 Sulfur globule protein CV2.
SQ SEQUENCE 124 AA; 12607 MW; 2220B28FFDC2525F CRC64;
```

```
Query Match 1.7%; Score 7; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 APVAPAA 107
DB 111 APVAPAA 117
|||||

RESULT 15
CU24_ARADI STANDARD; PRT; 126 AA.
AC P80516;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Adult-specific rigid cuticular protein 12.4 (ACP 12.4).
OS Araneus diadematus (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Araneidae; Araneus.
OX NCBI_TaxID=45920;
RN [1]
RP SEQUENCE.
RC TISSUE=Cuticle;
RX MEDLINE=97166616; PubMed=9014336;
RA Norup T., Berg T., Stenholm H., Andersen S.O., Hoejrup P.;
RT "Purification and characterization of five cuticular proteins from
the spider Araneus diadematus.";
RL Insect Biochem. Mol. Biol. 26:907-915(1996).
CC -1- FUNCTION: COMPONENT OF THE RIGID CUTICLE OF THE SPIDER.
CC -1- MASS SPECTROMETRY: MW=12420.5; METHOD=Electrospray.
CC -1- SIMILARITY: Contains 1 cuticle consensus domain.
DR InterPro; IPR000618; Insect cuticle.
DR Pfam; PF00379; Chitin bind_4; 1.
DR PROSITE; PS00233; CUTICLE; 1.
KW Structural protein; Cuticle.
FT DOMAIN 73 83 POLY-ALA.
SQ SEQUENCE 126 AA; 12421 MW; F9633E26A877F6DE CRC64;

Query Match 1.7%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 PAAPVAP 105
DB 92 PAAPVAP 98
|||||

RESULT 16
RISB_LACLA STANDARD; PRT; 154 AA.
AC Q9CGU6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 6,7-dimethyl-8-ribityllumazine synthase [EC 2.5.1.9] (DMRL synthase)
[umazine synthase] (Riboflavin synthase beta chain).
GN RISB OR LLO996.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarre K.,
Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -1- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
```

```

CC ribityl-amino-2,4-(1H,3H)-pyrimidinone and L-3,4-dihydroxy-2-
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
CC 2,4(1H,3H)-pyrimidinone with L-3,4-dihydroxy-2-butanone-4-
CC phosphate yielding 6,7-dimethyl-8-lumazine (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine =
CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC -!- PATHWAY: Riboflavin biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the DMRL synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE006333; AAK05094.1; -.
DR HSP; D86749; D86749.
DR HSP; P11998; IRVV.
DR HAMAP; MF_00178; -.
DR InterPro; IPR002180; DMRL synthase.
DR Pfam; PF00885; DMRL synthase; 1.
DR ProDom; PD003664; DMRL synthase; 1.
DR TIGRFAMs; TIGR00114; ribH; 1.
KW Riboflavin biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 154 AA; 16696 MW; 2E06CD2C2979A3BF CRC64;
-----
Query Match 1.7%; Score 7; DB 1; Length 154;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 273 SGALDAL 279
Db 32 SGALDAL 38
|||||
RESULT 17
RIB2 PHOLE STANDARD; PRT; 156 AA.
AC Q93E92;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
DE (Lumazine synthase) (Riboflavin synthase beta chain).
GN RIBH.
OS Photobacterium leiognathi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Photobacterium.
OX NCBI_TaxID=658;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=PL741;
RA MEDLINE=21290647; PubMed=11396941;
RA Lin J.-W., Chao Y.-F., Weng S.-F.;
RT "Riboflavin synthesis genes ribE, ribB, ribH, ribA reside in the lux
RL operon of Photobacterium leiognathi.";
RL Biochem. Biophys. Res. Commun. 284:587-595(2001).
CC -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
CC ribityl-amino-2,4(1H,3H)-pyrimidinone and L-3,4-dihydroxy-2-
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
CC 2,4(1H,3H)-pyrimidinone with L-3,4-dihydroxy-2-butanone-4-
CC phosphate yielding 6,7-dimethyl-8-lumazine (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine =
CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC -!- PATHWAY: Riboflavin biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the DMRL synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF364106; AAK83294.1; -.
DR HAMAP; MF_00178; -.
DR InterPro; IPR002180; DMRL synthase.
DR Pfam; PF00885; DMRL synthase; 1.
DR ProDom; PD003664; DMRL synthase; 1.
DR TIGRFAMs; TIGR00114; ribH; 1.
KW Riboflavin biosynthesis; Transferase.
SQ SEQUENCE 156 AA; 16423 MW; B53E2727C6BCD0A5 CRC64;
-----
Query Match 1.7%; Score 7; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 273 SGALDAL 279
Db 32 SGALDAL 38
|||||
RESULT 18
RRF TREPA STANDARD; PRT; 183 AA.
AC O83613;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome recycling factor (Ribosome releasing factor) (RRF).
DE FRR OR TP0604.
GN Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RA MEDLINE=98332770; PubMed=965876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Peterback T.,
RA McDonald L., Artiaach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -!- FUNCTION: RESPONSIBLE FOR THE RELEASE OF RIBOSOMES FROM MESSENGER
CC RNA AT THE TERMINATION OF PROTEIN BIOSYNTHESIS. MAY INCREASE THE
CC EFFICIENCY OF TRANSLATION BY RECYCLING RIBOSOMES FROM ONE ROUND OF
CC TRANSLATION TO ANOTHER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RRF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001235; AAC65577.1; -.
DR PIR; F71304; F71304.
DR HSP; Q9X1B9; 1DD5.
DR TIGR; TP0604; -.
DR HAMAP; MF_00040; -.
DR InterPro; IPR002661; RRF.
DR Pfam; PF01765; RRF; 1.

```

```

DR ProDom; PD004103; RRF; 1.
DR TIGRFAMs; TIGR00496; frf; 1.
KW Protein biosynthesis; Complete proteome.
SQ SEQUENCE 183 AA; 20503 MW; AE16B5894DDD7C4B CRC64;

Query Match      1.7%; Score 7; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 DASVADV 26
   |||||
Db 163 DASVADV 169

RESULT 19
KGUA_STRP3
ID KGUA_STRP3 STANDARD; PRT; 211 AA.
AC Q8P001;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Guanylate kinase (EC 2.7.4.8) (GMP kinase)
GN GMK OR SPYM3_1375 OR SPS0487 OR SPYM18_1842.
OS Streptococcus pyogenes (serotype M3), and
OC Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466, 186103;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.P., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=SSI-1 / Serotype M3;
RA Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yamashita A.,
RA Yamazaki K., Okahashi N., Kawabata S., Yasunaga T., Hattori M.,
RA Hayashi H., Hamada S.;
RT "The genome of invasive Streptococcus pyogenes; a comparative analysis
RT of S. pyogenes SSI-1, SF370 and MGAS8232.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
CC -1- FUNCTION: Essential for recycling GMP and indirectly, cGMP.
CC -1- CATALYTIC ACTIVITY: ATP + GMP = ADP + GDP.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE GUANYLATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE014162; AA079982.1; -
DR EMBL; AP005142; BAC63582.1; -
DR EMBL; AE010076; AAL98191.1; -
DR HAMAP; MF_00328; -; 1.
DR InterPro; IPR000619; Guanylate_kin_1.
DR Pfam; PF00625; Guanylate_kin; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS00552; GUANYLATE_KINASE_2; 1.
KW Transferase; Kinase; ATP-binding; Complete proteome.
FT NP_BIND 12 19 ATP (BY SIMILARITY)
SQ SEQUENCE 211 AA; 24186 MW; 296B236E67A925D4 CRC64;

Query Match      1.7%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 VKRIIET 32
   |||||
Db 179 VKRIIET 185

RESULT 20
KGUA_STRP3
ID KGUA_STRP3 STANDARD; PRT; 211 AA.
AC Q99YM5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Guanylate kinase (EC 2.7.4.8) (GMP kinase)
GN GMK OR SPY1632.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Perretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -1- FUNCTION: Essential for recycling GMP and indirectly, cGMP.
CC -1- CATALYTIC ACTIVITY: ATP + GMP = ADP + GDP.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE GUANYLATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE006594; AAK34402.1; -
DR HSSP; P15454; 1GKY.
DR HAMAP; MF_00328; -; 1.
DR InterPro; IPR000619; Guanylate_kin.
DR Pfam; PF00625; Guanylate_kin; 1.
DR SMART; SM00072; GUKC; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS00552; GUANYLATE_KINASE_2; 1.
KW Transferase; Kinase; ATP-binding; Complete proteome.
FT NP_BIND 12 19 ATP (BY SIMILARITY)
SQ SEQUENCE 211 AA; 24176 MW; 296B236E657925D4 CRC64;

Query Match      1.7%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



```

QY      26 VKRIIET 32
DB      179 VKRIIET 185

RESULT 21
OLP1_LYCES
ID OLP1_LYCES STANDARD; PRT; 252 AA.
AC Q41350;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Osmotin-like protein precursor.
OS Osmotin-like protein precursor.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. VF36;
RX MEDLINE=97128324; PubMed=8972917;
RA Chen R., Wang F., Smith A.G.;
RT "A flower-specific gene encoding an osmotin-like protein from
RT Lycopersicon esculentum.";
RL Gene 179:301-302(1996).
RN [2]
RP SEQUENCE OF 25-33.
RA Robertson D., Mitchell G.P., Gilroy J.S., Gerrish C., Bolwell G.P.,
RA Slabas A.R.;
RL Submitted (JAN-1997) to the SWISS-PROT data bank.
CC -!- SUBCELLULAR LOCATION: Cell wall (potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L76632; A841124.1; -.
CC FIR; JC5237; JC5237.
CC HSGP; P25871; LAUN.
CC InterPro; IPR001938; Thaumatin.
CC Pfam; PF00314; thaumatin; 1.
CC PRINTS; PR00347; THAUMATIN.
CC ProDom; PD003321; Thaumatin; 1.
CC SMART; SM00205; THN; 1.
CC PROSITE; PS00316; THAUMATIN; 1.
CC Cell wall; Signal.
CC SIGNAL 1 24
CC CHAIN 25 252 OSMOTIN-LIKE PROTEIN.
CC SEQUENCE 252 AA; 27265 MW; 36031901A4F6E190 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 252;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      92 APATIAQ 98
DB      113 APATIAQ 119

RESULT 22
HIS4_SYNY3
ID HIS4_SYNY3 STANDARD; PRT; 256 AA.
AC P74561;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]

```

```

DE imidazole-4-carboxamide isomerase (EC 5.3.1.16)
DE (Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide
DE isomerase).
GN HISA OR SLR0652.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kareko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpou S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RA "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-5-[(5-
CC phosphoribosylamino)methylideneamino]imidazole-4-carboxamide = 5-
CC [(5-phospho-1-deoxyribulos-1-ylamino)methylideneamino]-1-(5-
CC phosphoribosyl)imidazole-4-carboxamide
CC -!- PATHWAY: Histidine biosynthesis; fourth step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -!- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90916; BAA18668.1; -.
CC FIR; S76756; S76756.
CC HAMAP; MF 01014; -.
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR006063; His4.
CC InterPro; IPR006062; His_biosynth.
CC Pfam; PF00977; His_biosynth; 1.
CC TIGRFAMs; TIGR00007; TIGR00007; 1.
CC Isomerase; Histidine biosynthesis; Complete proteome.
CC SEQUENCE 256 AA; 27350 MW; 2759585808P7F439 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 256;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      276 LDALROL 282
DB      179 LDALROL 185

RESULT 23
TNR9_MOUSE
ID TNR9_MOUSE STANDARD; PRT; 256 AA.
AC P20334;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Tumor necrosis factor receptor superfamily member 9 precursor (4-1BB
DE ligand receptor) (T-cell antigen 4-1BB) (CD137 antigen).
GN TNFRSF9 OR ILA OR LY63 OR CD137 OR CD157.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89184547; PubMed=2784565;
RA Kwon B.S., Weissman S.M.;

```



```
RT "cDNA sequences of two inducible T-cell genes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:1963-1967(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c;  
RX MEDLINE=94179805; PubMed=8133039;  
RA Kwon B.S., Kozak C.A., Kim K.K., Pickard R.T.;  
RT "Genomic organization and chromosomal localization of the T-cell  
antigen 4-1BB.";  
RL J. Immunol. 152:2256-2262(1994).  
RN [3]  
RP CHARACTERIZATION, AND SEQUENCE OF 25-29.  
RX MEDLINE=93139510; PubMed=7678621;  
RA Pollok K.E., Kim Y.-J., Zhou Z., Hurtado J., Kin K.K., Pickard R.T.,  
RT Kwon B.S.;  
RT "Inducible T cell antigen 4-1BB. Analysis of expression and  
function.";  
RL J. Immunol. 150:771-781(1993).  
CC -|- FUNCTION: Receptor for TNFSF14/4-1BBL. Possibly active during T  
cell activation.  
CC -|- SUBUNIT: PRINCIPALLY AN HOMODIMER, BUT ALSO FOUND AS A MONOMER.  
CC ASSOCIATES WITH F56-LCK. Interacts with TRAF1, TRAF2 AND TRAF3 (By  
similarity).  
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -|- TISSUE SPECIFICITY: EXPRESSED ON THE SURFACE OF ACTIVATED T-CELLS.  
CC -|- INDUCTION: OPTIMAL BY PMA AND IONOMYCIN.  
CC -|- SIMILARITY: Contains 4 TNFR-Cys repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; J04492; AAA40167.1; -;  
DR EMBL; U02567; AAA93113.1; -;  
DR PIR; B32393; B32393.  
DR PDB; 1DQJ; 26-SEP-01.  
DR MGD; MGI:1101059; Tnfrsf9.  
DR InterPro; IPR001368; TNFR_c6.  
DR Pfam; PF00020; TNFR_c6; 1.  
DR SMART; SM00208; TNFR; 2.  
DR PROSITE; PS00652; TNFR_NGFR_1; 1.  
DR PROSITE; PS00500; TNFR_NGFR_2; FALSE NEG.  
KW Receptor; Transmembrane; Glycoprotein; Repeat; Signal; 3D-structure.  
FT SIGNAL 1 24  
FT CHAIN 25 256 TUMOR NECROSIS FACTOR RECEPTOR  
FT FT SUPERFAMILY MEMBER 9.  
FT DOMAIN 25 187 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 188 208 POTENTIAL.  
FT DOMAIN 209 256 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 17 45 TNFR-CYS 1.  
FT REPEAT 46 85 TNFR-CYS 2.  
FT REPEAT 86 117 TNFR-CYS 3.  
FT REPEAT 118 159 TNFR-CYS 4.  
FT DISULFID 28 37 BY SIMILARITY.  
FT DISULFID 31 44 BY SIMILARITY.  
FT DISULFID 47 61 BY SIMILARITY.  
FT DISULFID 64 77 BY SIMILARITY.  
FT DISULFID 67 85 BY SIMILARITY.  
FT DISULFID 87 93 BY SIMILARITY.  
FT DISULFID 98 105 BY SIMILARITY.  
FT DISULFID 101 116 BY SIMILARITY.  
FT DISULFID 119 133 BY SIMILARITY.  
FT DISULFID 139 158 BY SIMILARITY.  
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 256 AA; 27598 MW; 93A10D03C60813C4 CRC64;  
  
Query Match 1.7%; Score 7; DB 1; Length 256;  
Best Local Similarity 100.0%; Pred. No. 61;  
  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 332 PEGGPGG 338  
DB 175 PEGGPGG 181  
  
RESULT 24  
LIGE PSEPA  
ID LIGE PSEPA STANDARD; PRT; 280 AA.  
AC P27457;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Beta-etherase (Beta-aryl ether cleaving enzyme).  
GN LIGE  
OS Pseudomonas paucimobilis (Sphingomonas paucimobilis).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;  
OC Sphingomonadaceae; Sphingomonas.  
OX NCBI_TaxID=13689;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.  
RC STRAIN=SYK-6;  
RX MEDLINE=92078104; PubMed=1744051;  
RA Masai E., Katayama Y., Kawai S., Nishikawa S., Yamasaki M.,  
RA Morohoshi N.;  
RT "Cloning and sequencing of the gene for a Pseudomonas paucimobilis  
enzyme that cleaves beta-aryl ether.";  
RL J. Bacteriol. 173:7950-7955(1991).  
CC -|- FUNCTION: ABLE TO DEGRADE VARIOUS DIMERIC LIGNIN COMPOUNDS.  
CC CATALYZES THE UNIQUE AND REDUCTIVE CLEAVAGE OF ARYLGLYCEROL-  
BETA-ARYL ETHER.  
CC -|- SUBCELLULAR LOCATION: Inner membrane-associated.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M69107; AAA25878.1; -;  
DR EMBL; D11473; BAA02032.1; -;  
DR PIR; A43749; A43749.  
DR InterPro; IPR004046; GST_Cterm.  
DR InterPro; IPR004045; GST_Nterm.  
DR Pfam; PF00043; GST_C; 1.  
DR Pfam; PF02798; GST_N; 1.  
KW Hydrolase; Lignin degradation; Inner membrane.  
FT INIT MET 0  
FT SEQUENCE 280 AA; 31939 MW; 7A5705D2A29D9C5B CRC64;  
  
Query Match 1.7%; Score 7; DB 1; Length 280;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 107 ASVARTP 113  
DB 198 ASVARTP 204  
  
RESULT 25  
NAR4 HUMAN  
ID NAR4 HUMAN STANDARD; PRT; 314 AA.  
AC Q93070; Q9BZ50; Q9BZ51; Q9HB06;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Ecto-ADP-ribosyltransferase 4 precursor (EC 2.4.2.31) (NAD(P)(+)-  
arginine ADP-ribosyltransferase 4) (Mono (ADP-ribosyl)transferase 4)  
DE (Dombrock blood group carrier molecule).  
GN DO OR ART4 OR DOK1.
```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF373017; AAK53821.1; --  
 DR EMBL; AF373016; AAK53821.1; JOINED.  
 DR EMBL; AF374727; AAK53821.1; JOINED.  
 DR GO; GO:0016020; C:membrane; ISS.  
 DR GO; GO:0003956; F:NAD(P)+-arginine ADP-ribosyltransferase act. . .; ISS.  
 DR GO; GO:0006525; P:arginine metabolism; ISS.  
 DR GO; GO:0006471; P:protein amino acid ADP-ribosylation; ISS.  
 DR InterPro; IPR000768; ART\_family.  
 DR Pfam; PF01129; ART; 1.  
 DR PRINTS; PRO0970; RIBTRNSFRASE.  
 DR PROSITE; PS01291; ART; 1.  
 DR Transferase; Glycosyltransferase; Glycoprotein; NAD; GPI-anchor;  
 KW Signal.  
 FT SIGNAL 1 46 POTENTIAL.  
 FT CHAIN 47 285 ECTO-ADP-RIBOSYLTRANSFERASE 4.  
 FT PROPEP 286 314 REMOVED IN MATURE FORM (POTENTIAL).  
 FT ACT\_SITE 251 251 BY SIMILARITY.  
 FT LIPID 285 285 GPI-ANCHOR (POTENTIAL).  
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 314 AA; 35861 MW; 1CD76A650B87A423 CRC64;  
 Query Match 1.7%; Score 7; DB 1; Length 314;  
 Best Local Similarity 100.0%; Pred.No. 73;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 107 ASVARTP 113  
 DB 141 ASVARTP 147  
 RESULT 27  
 CAHC\_HORVU STANDARD; PRT; 324 AA.  
 ID CAHC\_HORVU STANDARD; PRT; 324 AA.  
 AC P40880;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Carbonic anhydrase, chloroplast precursor (EC 4.2.1.1) (Carbonate  
 DE dehydratase).  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95303975; PubMed=7784519;  
 RA Bracey M.H., Bartlett S.G.;  
 RT "Sequence of a cDNA encoding carbonic anhydrase from barley.";  
 RL Plant Physiol. 108:433-434(1995).  
 CC -1- FUNCTION: Reversible hydration of carbon dioxide.  
 CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.  
 CC -1- SUBUNIT: Homohexamer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Chloroplast stroma (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE PLANT AND PROKARYOTIC CARBONIC  
 CC ANHYDRASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L36959; AAC41656.1; --  
 DR PIR; T04478; T04478.  
 DR InterPro; IPR001765; Prok\_Coanhd.  
 DR Pfam; PF00484; Pro CA; 1.  
 DR PROSITE; PS00704; PROK\_CO2\_ANHYDRASE 1; 1.  
 DR PROSITE; PS00705; PROK\_CO2\_ANHYDRASE 2; 1.  
 KW Lyase; Zinc; Chloroplast; Transit peptide.  
 FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).  
 FT CHAIN ? 324 CARBONIC ANHYDRASE.  
 SQ SEQUENCE 324 AA; 35074 MW; 026F541956F06E75 CRC64;  
 Query Match 1.7%; Score 7; DB 1; Length 324;  
 Best Local Similarity 100.0%; Pred.No. 76;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 100 AAPVAPA 106  
 DB 115 AAPVAPA 121  
 RESULT 28  
 G3P\_HALN1 STANDARD; PRT; 335 AA.  
 ID G3P\_HALN1 STANDARD; PRT; 335 AA.  
 AC Q9HSS7;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.59) (GAPDH)  
 DE (NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase).  
 DE GAP OR GAPB OR VNG0095G.  
 GN Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).  
 OS Archaea; Euryarchaeota; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OX NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ebbhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +  
 CC NAD(P)(+) = 3-phospho-D-glyceroyl phosphate + NAD(P)H.  
 CC -1- PATHWAY: Second phase of glycolysis; first step.  
 CC -1- SUBUNIT: Homotetramer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: Belongs to the glyceraldehyde 3-phosphate  
 CC dehydrogenase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF004977; AAG18725.1; --  
 DR PIR; A84170; A84170.  
 DR HSSP; P10618; 1CF2.  
 DR HAMAP; MF\_00559; -; 1.  
 DR InterPro; IPR000173; GAP\_dhndrogenase.

DR InterPro; IPR006436; GAPDH-II\_archae.

DR Pfam; PF00044; Gpdh; 1.

DR Pfam; PF02800; Gpdh\_C; 1.

DR ProDom; PD007761; GAPDH-II\_archae; 1.

DR TIGRFAMs; TIGR01546; GAPDH-II\_archae; 1.

DR PROSITE; PS00071; GAPDH; 1.

KW Glycolysis; Oxidoreductase; NAD; NADP; Complete proteome.

FT BINDING 140 140 GLYCERALDEHYDE 3-PHOSPHATE (BY

FT SIMILARITY).

SQ SEQUENCE 335 AA; 35682 MW; CA10FD928E368CB1 CRC64;

Query Match

Best Local Similarity 1.7%; Score 7; DB 1; Length 335;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 DASVADV 26

DB 112 DASVADV 118

RESULT 29

GB0 GEOCY

ID\_GB0 GEOCY STANDARD; PRT; 358 AA.

AC Q9XZV3;

DT 15-SEP-2003 (Rel. 42, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Guanine nucleotide-binding protein G(O), alpha subunit.

OS Geodia cydonium (Sponge).

OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;

OC Asterozoa; Geodiidae; Geodia.

OX NCBI TaxID=6047;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98119577; PubMed=9459489;

RA Seack J., Kruse M., Mueller W.B.G.;

RT "Evolutionary analysis of G-proteins in early metazoans: cloning of

alpha- and beta-subunits from the sponge Geodia cydonium.";

RL Biochim. Biophys. Acta 1401:93-103(1998).

CC -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE

INVOLVED AS MODULATORS OR TRANSDUCERS IN VARIOUS TRANSMEMBRANE

SIGNALING SYSTEMS. THE G(O) PROTEIN FUNCTION IS NOT CLEAR.

CC -!- SUBUNIT: G proteins are composed of 3 units (alpha, beta and

gamma). The alpha chain contains the guanine nucleotide binding

site.

CC -!- SIMILARITY: BELONGS TO THE G-ALPHA FAMILY. SUBFAMILY 1

CC (G1/O/T/Z).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; Y14247; CAB43526.1; --

DR HSSP; P04896; LAZT.

DR InterPro; IPR001019; Gproteins\_alpha.

DR Pfam; PF00503; G-alpha; 1.

DR PRINTS; PR00318; GPROTEINA.

DR ProDom; PD000281; Gproteins\_alpha; 1.

DR SMART; SM00275; G\_alpha; 1.

KW GTP-binding; Transducer; Multigene family; ADP-ribosylation;

KW Myristate; Palmitate; Lipoprotein.

FT INIT MET 0 0 BY SIMILARITY.

FT LIPID 1 1 MYRISTATE (BY SIMILARITY).

FT LIPID 3 3 PALMITATE (BY SIMILARITY).

FT NP\_BIND 41 48 GTP (BY SIMILARITY).

FT NP\_BIND 202 206 GTP (BY SIMILARITY).

FT NP\_BIND 271 274 GTP (BY SIMILARITY).

FT MOD\_RES 180 180 ADP-RIBOSYL[1] (BY ACTION OF CTX) (BY

SIMILARITY).

FT MOD\_RES 180 180

SQ SEQUENCE 358 AA; 40933 MW; E3D6956089837D98 CRC64;

Query Match

Best Local Similarity 1.7%; Score 7; DB 1; Length 358;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 355 TPERRA 361

DB 7 TPERRA 13

RESULT 30

POS1 BOVIN

ID\_POS1 BOVIN STANDARD; PRT; 360 AA.

AC O97552;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE POU domain, class 5, transcription factor 1 (Octamer-binding

transcription factor 3) (Oct-3) (Oct-4).

GN POU5F1.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99225334; PubMed=10208969;

RA van Eijk M.J.T., van Rooijen M.A., Modina S., Scesi L., Folkers G.,

van Tol H.T.A., Bevers M.M., Fisher S.R., Lewin H.A., Shehu D.,

Galli C., de Vauzeix C., Trounson A.O., Mummery C.L., Gandolfi P.;

"Molecular cloning, genetic mapping, and developmental expression of

bovine POU5F1.";

RT Biol. Reprod. 60:1093-1103(1999).

CC -!- FUNCTION: Transcription factor that binds to the octamer motif

('ATTGCAAT'). Prime candidate for an early developmental control

gene (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.

CC CLASS-5 SUBFAMILY.

CC -!- SIMILARITY: Contains 1 homeobox domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AF022987; AAD01757.1; --

DR EMBL; AF022986; AAD01757.1; JOINED.

DR HSSP; P20263; 10CP.

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR000327; POU domain.

DR Pfam; PF00046; homeobox; 1.

DR Pfam; PF00157; pou; 1.

DR PRINTS; PR00028; POUDOMAIN.

DR ProDom; PD000010; Homeobox; 1.

DR ProDom; PD000583; POU domain; 1.

DR SMART; SM00389; HOX; 1.

DR SMART; SM00352; POU; 1.

DR PROSITE; PS00027; HOMEBOX 1; 1.

DR PROSITE; PS00071; HOMEBOX 2; 1.

DR PROSITE; PS00035; POU 1; 1.

DR PROSITE; PS00465; POU 2; 1.

KW Homeobox; DNA-binding; Transcription regulation; Nuclear protein.

FT DOMAIN 138 212 POU.

FT DNA\_BIND 230 289 HOMEBOX.

FT SEQUENCE 360 AA; 38289 MW; DF0A326BE6329F08 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 360;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 PGVVPGA 271  
DB 49 PGVVPGA 55

```
RESULT 31
INC2_ECOLI
ID INC2_ECOLI STANDARD; PRT; 364 AA.
AC P07673;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein incC.
GN INCC.
OS Escherichia coli.
OG Plasmid IncP-alpha RK2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86232584; PubMed=3520485;
RA Thomas C.M., Smith C.A.;
RT "The trfB region of broad host range plasmid RK2: the nucleotide
RT sequence reveals incC and key regulatory gene trfB/korA/korD as
RT overlapping genes.";
RL Nucleic Acids Res. 14:4453-4469 (1986).
RN [2]
RP SEQUENCE OF 330-364 FROM N.A.
RX MEDLINE=88015606; PubMed=3309894;
RA Theophilus B.D.B., Thomas C.M.;
RT "Nucleotide sequence of the transcriptional repressor gene korB which
RT plays a key role in regulation of the copy number of broad host range
RT plasmid RK2.";
RL Nucleic Acids Res. 15:7443-7450 (1987).
RN [3]
RP SEQUENCE OF 342-364 FROM N.A.
RX MEDLINE=88118923; PubMed=3430606;
RA Kornacki J.A., Balderes P.J., Figurski D.H.;
RT "Nucleotide sequence of korB, a replication control gene of broad
RT host-range plasmid RK2.";
RL J. Mol. Biol. 198:211-222 (1987).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS ENCODED BY THE TRFB OPERON;
CC IT IS INVOLVED IN PLASMID MAINTENANCE AND REPLICATION.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative initiation;
CC Comment=2 isoforms, Long (shown here) and Short/Small, are
CC produced by alternative initiation;
CC -1- SIMILARITY: BELONGS TO THE PARA FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X03962; CAA27595.1; --
EMBL; Y00446; CAA68502.1; --
EMBL; X06543; CAA29789.1; --
DR Pfam; PF00070; ATPase_Para.
DR Pfam; PF00991; Para; 1.
KW Plasmid; DNA replication; Alternative initiation.
FT CHAIN 1 364 PROTEIN INCC, ISOFORM LONG.
FT CHAIN 106 364 PROTEIN INCC, ISOFORM SHORT.
FT INIT_MET 106 106 FOR ISOFORM SHORT.
SQ SEQUENCE 364 AA; 38134 MW; 36241B743C53E0BE CRC64;
```

Query Match 1.7%; Score 7; DB 1; Length 364;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 PGAGSGA 275  
DB 19 PGAGSGA 25

```
RESULT 32
1A01_PONPY
ID 1A01_PONPY STANDARD; PRT; 365 AA.
AC P16211;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Class I histocompatibility antigen, A-1 alpha chain precursor.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90201944; PubMed=1690682;
RA Lawlor D.A., Warren E., Ward F.E., Parham P.;
RT "Comparison of Class I MHC alleles in humans and apes.";
RL Immunol. Rev. 113:147-185 (1990).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M30680; AAA88035.1; --
DR FIR; I84432; I84432.
DR HSP; Q95352; 1HKH.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001039; MHC_I.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF01129; MHC_I; 1.
DR PRINTS; PR01638; MHCCLASS1.
DR ProDom; PD000050; MHC_I; 1.
DR SMART; SM00407; IGC1_1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 365 CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 25 114 A-1 ALPHA CHAIN.
FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.
FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.
FT DOMAIN 309 332 CONNECTING PEPTIDE.
FT DOMAIN 333 365 CYTOPLASMIC TAIL.
FT DISULFID 125 188 BY SIMILARITY.
FT DISULFID 227 283 BY SIMILARITY.
FT CARBOHYD 110 110 N-LINKED (GLCNAC... ) (BY SIMILARITY).
SQ SEQUENCE 365 AA; 40658 MW; 11A5BC183009CF70 CRC64;
```

Query Match 1.7%; Score 7; DB 1; Length 365;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      147 YSQASN 153
      |||||
Db      344 YSQASN 350

RESULT 33
T2AL ACICA
ID T2AL ACICA STANDARD; PRT; 366 AA.
AC P24546;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Type II restriction enzyme AccI (EC 3.1.21.4) (Endonuclease AccI)
DE (R:AccI).
GN ACCIR.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_TaxID=471;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN=ATCC 49823;
RX MEDLINE=91345839; PubMed=1368703;
RA Kawakami B., Hilzheber C., Nagatomo M., Oka M.;
RT "Cloning and nucleotide sequences of the AccI restriction-modification
RT genes in Acinetobacter calcoaceticus."
RL Agric. Biol. Chem. 55:1553-1559(1991).
CC -!- FUNCTION: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE GTMKAC AND
CC CLEAVES AFTER T-2.
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give
CC specific double-stranded fragments with terminal 5'-phosphates.
CC -!- SUBUNIT: Homotrimer.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; D10671; BRA01522.1; -.
DR PIR; J00469; J00469.
DR REBASE; 18; AccI.
KW Restriction system; Hydrolase; Nuclease; Endonuclease.
SQ SEQUENCE 366 AA; 42495 MW; DB663B74351C415E CRC64;

Query Match 1.7%; Score 7; DB 1; Length 366;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      110 ARTPTQA 116
      |||||
Db      28 ARTPTQA 34

RESULT 34
SGK2 MOUSE
ID SGK2 MOUSE STANDARD; PRT; 367 AA.
AC Q9QZS5; Q9R0P6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Serine/threonine-protein kinase Sgk2 (EC 2.7.1.-)
DE (Serum/glucocorticoid regulated kinase 2).
GN SGK2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20018032; PubMed=10548550;

```

```

RA Kobayashi T., Deak M., Morrice N., Cohen P.;
RT "Characterization of the structure and regulation of two novel
RT isoforms of serum- and glucocorticoid-induced protein kinase.";
RL Biochem. J. 344:189-197(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Haegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldairelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schrim L.M., Kanapin A., Matsuda H.A., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazier K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempie C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wu Ming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Haeizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki K., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP FUNCTION: Involved in the activation of potassium channels (By
RP similarity).
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- ALTERNATIVE PRODUCTS:
CC -!- Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9QZS5-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9QZS5-2; Sequence=VSP_004933;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -!- CAUTION: Not regulated by serum or glucocorticoids.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF169033; AAF12756.1; -
CC EMBL; AK050009; BAC34031.1; -
CC EMBL; BC026549; AAB26549.1; -
CC HSSP; P05132; 1CTP.
CC MGI; 1351318; Sgk2.
CC InterPro; IPR000961; Pkinase C.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR002290; Ser_thr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00433; pkinase C; 1.
CC ProDom; PD000001; Prot_kinase; 1.
CC SMART; SM00133; S_TK_X; 1.
CC SMART; SM00220; S_TKG; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00111; PROTEIN KINASE DOM; 1.
CC PROSITE; PS00108; PROTEIN KINASE ST; 1.
CC Transferrase; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation; Alternative splicing
KW DOMAIN 35 292
FT NP_BIND 41 49 ATP (BY SIMILARITY).
FT BINDING 64 64 ATP (BY SIMILARITY).
FT ACT_SITE 159 159 BY SIMILARITY.
FT MOD_RES 193 193 PHOSPHORYLATION (BY PDPK1)
FT VARSPIC 171 199 Missing (in isoform 2).
FT CONFLICT 77 77 MISSING (IN REF. 3).
FT SEQUENCE 367 AA; 41359 MW; 668C04B1A1E9E33A CRC64;

Query Match 1.7%; Score 7; DB 1; Length 367;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 ASSSGAS 85
DB 343 ASSSGAS 349
|||||||

RESULT 35
ADHP RABIT
ID ADHP RABIT STANDARD; PRT; 378 AA.
AC O46649;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase class II isozyme 1 (EC 1.1.1.1).
GN ADH2-1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Liver;
RX MEDLINE=98151253; PubMed=9492289;
RA Svensson S., Hedberg J., Hoesog J.-O.;
RT "Structural and functional divergence of class II alcohol
RT dehydrogenase -- cloning and characterisation of rabbit liver isoforms
RT of the enzyme.";
RL Eur. J. Biochem. 251:236-243(1998).
CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).

```

```

CC -----
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family. Class-II subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ002388; CAA05362.1; -
CC HSSP; P00325; 1DEH.
CC InterPro; IPR002328; ADH_zinc.
CC InterPro; IPR002085; Adh_zn_family.
CC Pfam; PF00107; ADH_zinc_N; 1.
CC PROSITE; PS00059; ADH_ZINC; 1.
CC Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
FT INIT_MET 0 0
FT METAL 46 46 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 68 68 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 98 98 ZINC 2 (BY SIMILARITY).
FT METAL 101 101 ZINC 2 (BY SIMILARITY).
FT METAL 104 104 ZINC 2 (BY SIMILARITY).
FT METAL 112 112 ZINC 2 (BY SIMILARITY).
FT METAL 175 175 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 378 AA; 40105 MW; EA21E1EF5920FD CRC64;

Query Match 1.7%; Score 7; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TLKGTFN 13
DB 318 TLKGTFN 324
|||||||

RESULT 36
SUCC CHLTR
ID SUCC CHLTR STANDARD; PRT; 386 AA.
AC O84828;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Succinyl-CoA synthetase beta chain (EC 6.2.1.5) (SCS-beta).
GN SUCC OR CT821.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/OW-3/Cx;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -!- CATALYTIC ACTIVITY: ATP + succinate + CoA = ADP + succinyl-CoA +
CC phosphate.
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SUBUNIT: Composed of an alpha chain and a beta chain (By
CC similarity).
CC -!- SIMILARITY: Belongs to the succinate/malate CoA ligase beta
CC subunit family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
CC
CC
DR EMBL; AB001354; AAC68418.1; -.
DR PIR; G71467; G71467.
DR HSP; P07460; 2SCU.
DR HAMAP; MF_00558; -. 1.
DR InterPro; IPR003135; ATP-grasp.
DR InterPro; IPR005809; CoA_lig_beta.
DR InterPro; IPR005811; CoA_ligase.
DR Pfam; PF02222; ATP-grasp; 1.
DR Pfam; PF00549; ligase-CoA; 1.
DR TIGRfams; TIGR01016; sucCoabeta; 1.
DR PROSITE; PS01217; SUCCINYL-CoA_lig_3; 1.
KW Ligase; Tricarboxylic acid cycle; Complete proteome.
SQ SEQUENCE 386 AA; 41863 MW; A74FAAC97FEB9309 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 386;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 71 VIMLSKA 77.
Db 122 VIMLSKA 128

RESULT 37
EFTU_XYLFA
ID EFTU_XYLFA STANDARD; PRT; 395 AA.
AC Q9PQ9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Elongation factor Tu (EF-Tu).
GN (TUPA OR XF2628) AND (TUF6 OR XF2640).
OS Xylolla fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylolla.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-9a5c;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvares A.R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Canargo A.A., Canargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylolla fastidiosa."
RL Nature 406:151-159(2000).
RN [2]
RP SEQUENCE OF 1-12.

```

```

RC STRAIN-9a5c;
RA Smolka M., Winck F., Martins D., Santoro C., Brum I., Galembeck E.,
RA Filho H., Machado M., Lemos E., Toyama M., Marangoni S.,
RA Novello J.C.;
RT "Proteomics of the plant pathogen Xylolla fastidiosa strain 9a5c:
RT implications for the study of pathogenesis."
RL Unpublished observations (JAN-2002).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).


-----



EMBL; AB004069; AAF85425.1; -.
EMBL; AB004071; AAF85437.1; -.
PIR; A82532; A82532.
HSSP; P07157; 1AIP.
HAMAP; MF_00118; -. 1.
InterPro; IPR004541; EF-Tu.
InterPro; IPR000795; EF_Gtpbind.
InterPro; IPR004160; EFTU_Cterm.
InterPro; IPR004161; EFTU_D2.
InterPro; IPR005225; Small_GTP.
Pfam; PF00009; GTP_EFTU; 1.
Pfam; PF03144; GTP_EFTU_D2; 1.
Pfam; PF03143; GTP_EFTU_D3; 1.
PRINTS; PR00315; ELONGATNFCT.
TIGRfams; TIGR00485; EF-Tu; 1.
TIGRfams; TIGR00231; small_GTP; 1.
PROSITE; PS00301; EFACITOR_GTP; 1.
Elongation factor; Protein biosynthesis; GTP-binding;
Complete proteome.
FT INIT MET 0
FT NP_BIND 18 25 GTP (BY SIMILARITY).
FT NP_BIND 80 84 GTP (BY SIMILARITY).
FT NP_BIND 135 138 GTP (BY SIMILARITY).
SQ SEQUENCE 395 AA; 42745 MW; 015C6DFD9B179AC0 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 395;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 NPERAID 196
Db 203 NPERAID 209

RESULT 38
FXQ1_MOUSE
ID FXQ1_MOUSE STANDARD; PRT; 400 AA.
AC O70220;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Forkhead box protein Q1 (Hepatocyte nuclear factor 3 forkhead homolog
DE 1) (HNF-3/forkhead-like protein 1) (HFH-1) (HFH-11).
GN FOXQ1 OR HFH1 OR HFH1L.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.


```



```

RC STRAIN=129; TISSUE=Kidney;
RX MEDLINE=98392851; PubMed=9726250;
RA Frank S., Zoll B.;
RT "Mouse HNF-3/fork head homolog-1-like gene: structure, chromosomal
RT location, and expression in adult and embryonic kidney.";
RL DNA Cell Biol. 17:679-688 (1998).
RN [2]
RP REVISIONS.
RA Pasche B., Bieller A., Zoll B.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP DISEASE.
RX MEDLINE=21207067; PubMed=11309849;
RA Hong H.-K., Noveroske J.K., Headon D.J., Liu T., Sy M.S.,
RA Justice M.J., Chakravarti A.;
RT "The winged helix/forkhead transcription factor Foxq1 regulates
RT differentiation of hair in statin mice.";
RL Genesis 29:163-171 (2001).
CC CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC CC -1- TISSUE SPECIFICITY: Expressed in kidney and stomach. Expression in
CC CC the outer medulla of the kidney and the transitional epithelium.
CC CC -1- DISEASE: Defects in FOXQ1 are the cause of the statin (sa)
CC CC phenotype and results in a silky coat with high sheen arising from
CC CC structurally abnormal medulla cells and defects in differentiation
CC CC of the hair shaft.
CC CC -1- SIMILARITY: Contains 1 fork-head domain.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL; AF010405; AAC12973.2; -.
DR HSSP; Q63245; 2HFH.
DR MGP; MGI:1298228; Foxq1.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DNA BIND 115 210 FORK-HEAD.
SQ SEQUENCE 400 AA; 41342 MW; E6C3B5F53F3BC42B CRC64;

Query Match 1.7k; Score 7; DB 1; Length 400;
Best Local Similarity 100.0k; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 QQAAPAA 136
Db 236 QQAAPAA 242
|||||
|||||

RESULT 39
PROV_ECOLI
ID PROV_ECOLI STANDARD; PRT; 400 AA.
AC P14175;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycine betaine/L-proline transport ATP-binding protein prov.
GN PROV OR B2677.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=89197759; PubMed=2649479;
RA Gowrishankar J.;
RT "Nucleotide sequence of the osmoregulatory proU operon of Escherichia
RT coli.";
RL J. Bacteriol. 171:1923-1931 (1989).
RN [2]
RP ERRATUM.
RA Gowrishankar J.;
RL J. Bacteriol. 172:1165-1165 (1990).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MGI1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=K12; PubMed=9205837;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba H., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oahina T., Oyama S., Saito N., Sampa G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113 (1997).
RN [5]
RP SEQUENCE OF 1-224 FROM N.A.
RX STRAIN=K12;
RX MEDLINE=90113884; PubMed=2691838;
RA Stirling D.A., Hulton C.S.J., Waddell L., Park S.F., Stewart G.S.A.B.,
RA Booth I.R., Higgins C.F.;
RT "Molecular characterization of the proU loci of Salmonella
RT typhimurium and Escherichia coli encoding osmoregulated glycine
RT betaine transport systems.";
RL Microbiol. 3:1025-1038 (1989).
CC CC -1- FUNCTION: INVOLVED IN A MULTICOMPONENT BINDING-PROTEIN-DEPENDENT
CC CC TRANSPORT SYSTEM FOR GLYCINE BETAINES/L-PROLINE.
CC CC -1- SUBCELLULAR LOCATION: Inner membrane-associated.
CC CC -1- SIMILARITY: Belongs to the ABC transporter family.
CC CC -1- SIMILARITY: Contains 2 CBS domains.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL; M24856; AAA24427.1; -.
DR EMBL; AE000352; AAC75724.1; -.
DR EMBL; D90891; BAA16542.1; -.
DR EMBL; X52694; CAA36923.1; -.
DR PIR; JS0128; BVECPV.
DR HSP; Q58663; IG5H.
DR ECO2DBASE; G033.5; 6TH EDITION.
DR EcoGene; EG10771; prov.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR000644; CBS_domain.
DR InterPro; IPR005892; Prov.
DR Pfam; PF00005; ABC_tran; 1.
DR Pfam; PF00571; CBS; 1.
DR ProDom; PD000006; ABC_transporter; 1.

```

```

DR SMART; SM00382; AAA; 1.
DR SMART; SM00116; CBS; 1.
DR TIGRFAMS; TIGR01186; PROV; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Inner membrane; Amino-acid transport; Transport; ATP-binding; Repeat;
KW CBS domain; Complete proteome.
FT DOMAIN 29 265 ABC_TRANSPORTER.
FT DOMAIN 279 334 CBS 1.
FT DOMAIN 340 393 CBS 2.
FT NP BIND 61 68 ATP (BY SIMILARITY).
FT NP BIND 61 68 ATP (BY SIMILARITY).
SQ SEQUENCE 400 AA; 44162 MW; 45A98D45F028B1B9 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 400;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 275 ALDALRQ 281
Db 145 ALDALRQ 151

RESULT 40
PROV_SALTY STANDARD; PRT; 400 AA.
AC P17328;
DT 01-AUG-1990 (Rel. 15, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycine betaine/L-proline transport ATP-binding protein prov.
GN PROV OR STM2809.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=90113884; PubMed=2691838;
RA Stirling D.A., Hulton C.S.J., Waddell L., Park S.F., Stewart G.S.A.B.,
RA Booth I.R., Higgins C.F.;
RT "Molecular characterization of the proU loci of Salmonella
RT typhimurium and Escherichia coli encoding osmoregulated glycine
RT betaine transport systems."
RL Mol. Microbiol. 3:1025-1038(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE OF 1-291 FROM N.A.
RX MEDLINE=89359099; PubMed=2548994;
RA "Nucleotide sequence of the transcriptional control region of the
RT osmotically regulated proU operon of Salmonella typhimurium and
RT identification of the 5' endpoint of the proU mRNA."
RL J. Bacteriol. 171:4694-4706(1989).
RN [4]
RP SEQUENCE OF 1-196 FROM N.A.
RX MEDLINE=94253018; PubMed=8195103;
RA Jordan A., Gilbert I., Barbe J.;
RT "Cloning and sequencing of the genes from Salmonella typhimurium
RT encoding a new bacterial ribonucleotide reductase."
RL J. Bacteriol. 176:3420-3427(1994).
RN [5]
RP FUNCTION: INVOLVED IN A MULTICOMPONENT BINDING-PROTEIN-DEPENDENT
CC

```

TRANSPORT SYSTEM FOR GLYCINE BETAINE/L-PROLINE.

- SUBCELLULAR LOCATION: Inner membrane-associated.

- SIMILARITY: Belongs to the ABC transporter family.

- SIMILARITY: Contains 2 CBS domains.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----

EMBL; X52693; CAA36921.1; --

EMBL; AE008828; AAL21694.1; --

EMBL; M26063; AAB8621.1; --

EMBL; X73226; CAA51696.1; --

PIR; S05374; QREBVT.

HSSP; Q58663; 1G6H.

StyGene; SG10312; prov.

InterPro; IPR003593; AAA\_ATPase.

InterPro; IPR003439; ABC\_transporter.

InterPro; IPR000644; CBS\_domain.

InterPro; IPR005892; Prov.

Pfam; PF00005; ABC\_tran; 1.

Pfam; PF00571; CBS; 2.

ProDom; PD000006; ABC\_transporter; 1.

SMART; SM00382; AAA; 1.

SMART; SM00116; CBS; 1.

TIGRFAMS; TIGR01186; prov; 1.

PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.

PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.

KW Inner membrane; Amino-acid transport; Transport; ATP-binding; Repeat;

KW CBS domain; Complete proteome.

FT DOMAIN 29 265 ABC\_TRANSPORTER.

FT DOMAIN 279 334 CBS 1.

FT NP BIND 61 68 ATP (BY SIMILARITY).

FT NP BIND 61 68 ATP (BY SIMILARITY).

FT CONFLICT 158 158 A -> R (IN REF. 1).

SQ SEQUENCE 400 AA; 44124 MW; FDF58DE0A9A792C2 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 400;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 275 ALDALRQ 281

Db 145 ALDALRQ 151

RESULT 41

SGK2\_HUMAN

ID SGK2\_HUMAN STANDARD; PRT; 427 AA.

AC Q9HBY6; Q9UKG6;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Serine/threonine-protein kinase Sgk2 (EC 2.7.1.37)

DE (Serum/glucocorticoid regulated kinase 2).

GN SGK2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1 AND 2), PHOSPHORYLATION ON THR-253, AND

RP MUTAGENESIS OF SER-416.

RX MEDLINE=20018032; PubMed=10548550;

RA Kobayashi T., Deak M., Morrice N., Cohen P.;

RT "Characterization of the structure and regulation of two novel

RT isoforms of serum- and glucocorticoid-induced protein kinase.";

RL Biochem. J. 344:189-197(1999).

RN [2]

SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,  
 RA Ellington A.G., Griffiths J.A., Frazer A.A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S.A., Harley J.L., Heath P.D., Ho S.E., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Leheslahti M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConnachie L.J., McVay K., McMurray A.D.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Stuke C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracner A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Colon;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marasina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwen P.C., Kerkman K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.J., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Pahay J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP FUNCTION.  
 RX MEDLINE=22284526; PubMed=12397388;  
 RA Gamber N., Fillon S., Peng Y., Friedrich B., Lang P.A., Henke G.,  
 RA Huber S.M., Kobayashi T., Cohen P., Lang P.;  
 RT "K(+) channel activation by all three isoforms of serum- and  
 glucocorticoid-dependent protein kinase SGK.";  
 RL Pflugers Arch. 445:60-66(2002).  
 CC -1- FUNCTION: Involved in the activation of potassium channels.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=2; Synonyms=beta;  
 CC IsoId=Q9HBY8-1; Sequence=Displayed;  
 CC Name=1; Synonyms=alpha;  
 CC IsoId=Q9HBY8-2; Sequence=VSP\_004932;  
 CC -1- TISSUE SPECIFICITY: Highly expressed in liver, kidney and  
 CC pancreas, and at lower levels in brain.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -1- CAUTION: Not regulated by serum or glucocorticoids.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 DR EMBL; AF186470; AAG17012.1; -;  
 DR EMBL; AF169034; AAF12757.2; -;  
 DR EMBL; Z98752; CAC18509.1; -;  
 DR EMBL; BC014037; AAH14037.1; -;  
 DR HSSP; P05132; 1ATP.  
 DR Genew; HGNC:13900; SGK2.  
 DR MIM; 607589; -;  
 DR InterPro; IPR000961; Pkinase C.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00433; Pkinase C; 1.  
 DR ProDom; PD0000001; Prot\_kinase; 1.  
 DR SMART; SM00133; S\_TK\_X; 1.  
 DR SMART; SM00220; S\_TK; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Phosphorylation; Alternative splicing.  
 FT DOMAIN 95 352  
 FT NP\_BIND 101 109  
 FT BINDING 124 124  
 FT ACT\_SITE 219 219  
 FT MOD\_RES 253 253  
 FT VARSELIC 1 60  
 FT MUTAGEN 416 416 S->D: INCREASED ACTIVATION.  
 FT SEQUENCE 427 AA; 47604 MW; D8F0FA6DF54B1370 CRC64;  
 SQ  
 Query Match 1.7%; Score 7; DB 1; Length 427;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 79 ASSSGAS 85  
 DB 403 ASSSGAS 409  
 |||||  
 RESULT 42  
 YAP1\_CHICK  
 ID\_YAP1\_CHICK STANDARD; PRT; 448 AA.  
 AC P46936;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 65 kDa Yes-associated protein (YAP65).  
 GN YAP1 OR YAP65.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]\_TaxID=9031;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=White leghorn;  
 RX MEDLINE=94309887; PubMed=8035999;  
 RA Sudol M.;  
 RT "Yes-associated protein (YAP65) is a proline-rich phosphoprotein that  
 RT binds to the SH3 domain of the Yes proto-oncogene product.";  
 RL Oncogene 9:2145-2152(1994).  
 CC -1- FUNCTION: BINDS TO THE SH3 DOMAIN OF THE YES KINASE. ALSO BINDS TO  
 CC OTHER SIGNALING MOLECULES THAT CONTAIN SH3 DOMAINS INCLUDING NCK,  
 CC CRK AND SRC.

```

CC -1- PTM: PHOSPHORYLATED ON SERINE RESIDUES.
CC -1- SIMILARITY: Contains 1 WW domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76483; CAA54021.1; -.
DR PIR; I50730; I50730.
DR InterPro; IPR005153; M0CH.
DR InterPro; IPR002349; WW.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00397; WW; 1.
DR PRINTS; PR00403; WWDOMAIN.
DR SMART; SM00456; WW; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
KW Phosphorylation.
FT DOMAIN 169 202 WW
SQ SEQUENCE 448 AA; 47822 MW; 719CC8D0F879A38D CRC64;

Query Match 1.7%; Score 7; DB 1; Length 448;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 PGAGSGA 275
DB 28 PGAGSGA 34
|||||

RESULT 43
ANT3 MOUSE
ID ANT3_MOUSE STANDARD; PRT; 465 AA.
AC P32261;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antithrombin-III precursor (ATIII).
GN SERPINC1 OR AT3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93069082; PubMed=1440494;
RA Wu J.K., Sheffield W.P., Blajchman M.A.;
RT "Molecular cloning and cell-free expression of mouse antithrombin
RT III."
RL Thromb. Haemost. 68:291-296(1992).
CC -1- FUNCTION: MOST IMPORTANT SERINE PROTEASE INHIBITOR IN PLASMA
CC THAT REGULATES THE BLOOD COAGULATION CASCADE. AT-III INHIBITS
CC THROMBIN AS WELL AS FACTORS IXA, XA AND XIA. ITS INHIBITORY
CC ACTIVITY IS GREATLY ENHANCED IN THE PRESENCE OF HEPARIN.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Plasma.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S47225; AAB23965.1; -.
DR HSSP; P01008; IATH.
DR MGD; MGI:88095; Serpinc1.

```

```

DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR SMART; SM00093; SERPIN; 1.
DR PROSITE; PS00284; SERPIN; 1.
KW Serpin; Serine protease inhibitor; Glycoprotein; Heparin-binding;
KW Plasma; Blood coagulation; Signal.
FT CHAIN 1 32 BY SIMILARITY.
FT BINDING 33 465 ANTITHROMBIN-III.
FT BINDING 82 82 HEPARIN (BY SIMILARITY).
FT BINDING 162 162 HEPARIN (BY SIMILARITY).
FT BINDING 178 178 HEPARIN (BY SIMILARITY).
FT ACT_SITE 426 427 REACTIVE BOND.
FT DISULFID 41 161 BY SIMILARITY.
FT DISULFID 54 128 BY SIMILARITY.
FT DISULFID 280 463 BY SIMILARITY.
FT CARBOHYD 129 129 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 465 AA; 52003 MW; 5CE087E98874E35D CRC64;

Query Match 1.7%; Score 7; DB 1; Length 465;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 PGAGSGA 275
DB 4 PGAGSGA 10
|||||

RESULT 44
FXD3 MOUSE
ID FXD3_MOUSE STANDARD; PRT; 465 AA.
AC G61050;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Forkhead box protein D3 (HNF3/FH transcription factor genesis)
DE Hepatocyte nuclear factor 3 forkhead homolog 2) (HFH-2).
GN FOXD3 OR HPH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Hromas R.A., Costa R.H., Xu D., Sutton J.L.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Labosky P.A., Kaestner K.H.;
RT "The winged helix transcription factor Hfh2 is expressed in neural
RT crest and spinal cord during mouse development."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 1 fork-head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U41047; AAA87569.1; -.
DR EMBL; AF067421; AAC28352.1; -.
DR HSSP; Q63245; 2HFH.
DR TRANSFAC; T04166; -.
DR MGD; MGI:1347473; Foxd3.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.

```

```

DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 106 115 POLY-GLY.
FT DNA_BIND 131 225 FORK-HEAD.
FT DOMAIN 252 257 POLY-ALA.
FT DOMAIN 265 270 POLY-ALA.
FT DOMAIN 275 281 POLY-ALA.
FT DOMAIN 380 399 POLY-GLY.
FT DOMAIN 447 457 POLY-ALA.
SQ SEQUENCE 465 AA; 47092 MW; 6F8B5B3D8C7564D CRC64;

Query Match 1.7%; Score 7; DB 1; Length 465;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 PGAGSGA 275
DB 110 PGAGSGA 116
|||||

RESULT 45
CO1B_MOUSE STANDARD; PRT; 484 AA.
AC Q9WUJ3; Q9CVA2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Coronin 1B (Coronin 2).
GN CORO1B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=98449467; PubMed=9778037;
RA Okumura M., Kung C., Wong S., Rodgers M., Thomas M.L.;
RT "Definition of family of coronin-related proteins conserved between
RT humans and mice: close genetic linkage between coronin-2 and CD45-
RT associated protein."
RL DNA Cell Biol. 17:779-787(1998).
RN 2
RP SEQUENCE OF 163-484 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Stomach;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Peole G., Quackenbush J.,
RA Schirnl L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: MAY BE INVOLVED IN CYTOKINESIS, MOTILITY, AND SIGNAL
CC TRANSDUCTION (BY SIMILARITY).
CC -!- SUBUNIT: BINDS ACTIN (BY SIMILARITY).

```

---

```

CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Contains 5 WD repeats.
CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF143956; AAD32704.1; -.
CC DR EMBL; AK008947; BAB25985.1; -.
CC DR MGD; MGI:1345963; Corolb.
CC DR InterPro; IPR001680; WD40.
CC DR Pfam; PF00400; WD40; 3.
CC DR ProDom; PD000018; WD40; 1.
CC DR SMART; SM00320; WD40; 3.
CC DR PROSITE; PS00678; WD_REPEATS_1; 1.
CC PROSITE; PS00082; WD_REPEATS_2; 2.
CC PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 80 120 WD 1.
FT REPEAT 130 170 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 217 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 444 482 COILED COIL (POTENTIAL).
FT CONFLICT 393 393 R -> G (IN REF. 2).
SQ SEQUENCE 484 AA; 53912 MW; 9631CC02E7EAC72F CRC64;

Query Match 1.7%; Score 7; DB 1; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 SGASTAT 88
DB 423 SGASTAT 429
|||||

RESULT 46
GALT_LACCA STANDARD; PRT; 486 AA.
ID -GALT_LACCA
AC O84904;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Galactose-1-phosphate uridylyltransferase (EC 2.7.7.12) (Gal-1-P
DE uridylyltransferase) (UDP-glucose--hexose-1-phosphate
DE uridylyltransferase).
GN GALT.
OS Lactobacillus casei.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1582;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=64H;
RX MEDLINE=98268934; PubMed=9603808;
RA Bettenbrock K., Alpert C.-A.;
RT "The gal genes for the leloir pathway of Lactobacillus casei 64H.";
RL Appl. Environ. Microbiol. 64:2013-2019(1998).
CC -!- CATALYTIC ACTIVITY: UDP-glucose + alpha-D-galactose 1-phosphate =
CC alpha-D-glucose 1-phosphate + UDP-galactose.
CC -!- PATHWAY: Galactose metabolism; second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the galactose-1-phosphate
CC uridylyltransferase family 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF005933; AAC19330.1; -;  
DR HAWAP; MF\_00571; -; 1.  
DR InterPro; IPR000766; Galp transf II.  
DR InterPro; IPR005850; Galp Utransf C.  
DR InterPro; IPR005849; Galp\_Utransf\_N.  
DR InterPro; IPR005934; GalT2.  
DR Pfam; PF02744; Galp\_UDP\_tr.C; 1.  
DR Pfam; PF01087; Galp\_UDP\_transf; 1.  
DR TIGRFAMs; TIGR01239; GalT\_2; 1.  
DR PROSITE; PS01163; GAL P UDP TRANSF II; 1.  
KW Transferase; Nucleotidyltransferase; Galactose metabolism.  
SQ SEQUENCE 486 AA; 54161 MW; 91A8EDE4F449B035 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 486;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 PENVEAQ 209  
Db 444 PENVEAQ 450

## RESULT 47

Y222\_CABEL STANDARD; PRT; 498 AA.  
ID Y222\_CABEL Q21339;  
AC Q21339;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein K08E5.2 in chromosome III.  
GN K08E5.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Kershaw J.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RA Durbin R.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: BELONGS TO THE SLC13A FAMILY OF TRANSPORTERS.  
CC NADC SUBFAMILY.  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Z30974; CAA83225.2; -;  
DR WormPep; K08E5.2; CE28598.  
DR InterPro; IPR001898; Na/sul\_sympot.  
DR Pfam; PF00939; Na sulph sym; 1.  
DR TIGRFAMs; TIGR00785; Gass; 1.  
DR PROSITE; PS01271; NA SULFATE; 1.  
KW Hypothetical protein; Transmembrane; Transport.  
FT TRANSMEM 24 44 POTENTIAL.  
FT TRANSMEM 94 114 POTENTIAL.  
FT TRANSMEM 151 171 POTENTIAL.  
FT TRANSMEM 200 220 POTENTIAL.  
FT TRANSMEM 261 281 POTENTIAL.

FT TRANSMEM 284 304 POTENTIAL.  
FT TRANSMEM 332 352 POTENTIAL.  
FT TRANSMEM 371 391 POTENTIAL.  
FT TRANSMEM 428 448 POTENTIAL.  
SQ SEQUENCE 498 AA; 54517 MW; FAEDB2FAA694BCB1 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 498;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 264 NPGVVPQ 270  
Db 270 NPGVVPQ 276

## RESULT 48

DHAL\_MOUSE STANDARD; PRT; 500 AA.  
ID DHAL\_MOUSE AC P24549;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Aldehyde dehydrogenase 1A1 (EC 1.2.1.3) (Aldehyde dehydrogenase,  
DE cytosolic) (ALDH class 1) (ALDH1) (ALDH-E1).  
GN ALDH1A1 OR ALDH1 OR AHD2 OR AHD-2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c, and C57BL/6; TISSUE=Liver;  
RX MEDLINE=91276281; PubMed=2055490;  
RA Rongoparut P., Weaver S.;  
RT "Isolation and characterization of a cytosolic aldehyde  
RT dehydrogenase-encoding cDNA from mouse liver.";  
RL Gene 101:261-265(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c, C57BL/6J, and 129/REJ; TISSUE=Liver;  
RX MEDLINE=95085815; PubMed=7993664;  
RA Bond S.L., Singh S.M.;  
RT "DNA sequence analysis of the cytosolic acetaldehyde dehydrogenase  
RT gene (Ahd-2) in mouse strains with variable ethanol preferences.";  
RL Biochem. Med. Metab. Biol. 52:155-159(1994).  
CC -!- FUNCTION: IN ADDITION TO THE ACTIVITY ON ACETALDEHYDE AND RELATED  
CC SUBSTRATES, IS ALSO INVOLVED IN THE OXIDATION OF ALDEHYDES DERIVED  
CC FROM BIOGENIC AMINES SUCH AS EPINEPHRINE AND NOREPINEPHRINE, AS  
CC WELL AS THE ALDEHYDES GENERATED VIA LIPID PEROXIDATION. BINDS FREE  
CC RETINAL AND CELLULAR RETINOL-BINDING PROTEIN-BOUND RETINAL. CAN  
CC CONVERT/OXIDIZE RETINALDEHYDE TO RETINOIC ACID (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.  
CC -!- PATHWAY: Ethanol utilization; second step.  
CC -!- SUBUNIT: Homodimer.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE LIVER, LUNG, AND TESTES. IS  
CC APPARENTLY NOT EXPRESSED AT DETECTABLE LEVELS IN KIDNEY, STOMACH,  
CC OVARY, HEART, AND BRAIN.  
CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M74570; AAA37202.1; -;  
DR EMBL; M74571; AAA37203.1; -;  
DR EMBL; S75713; AAB32754.2; -;  
DR EMBL; S77047; -; NOT ANNOTATED\_CDS.  
DR PIR; JQ1004; JQ1004.

```

DR HSP; P51977; 1BX5.
DR SWISS-2DPAGE; P24549; MOUSE.
DR MGD; MGI:1353450; Aldh1a1.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Acetylation.
FT INIT MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 268 268 BY SIMILARITY.
FT ACT_SITE 302 302 BY SIMILARITY.
FT BINDING 299 304 ANTABUSE.
FT CONFLICT 7 7 A -> R (IN REF. 1).
FT CONFLICT 44 44 T -> S (IN REF. 2).
FT CONFLICT 50 50 H -> Q (IN REF. 2).
FT CONFLICT 86 86 R -> C (IN REF. 1).
FT CONFLICT 457 457 I -> M (IN REF. 2).
SQ SEQUENCE 500 AA; 54318 MW; 3E428154E7214B54 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 SPAQPAV 233
DB 2 SPAQPAV 8
|||||

RESULT 49
DHAI RAT
ID DHAI RAT STANDARD; PRT; 500 AA.
AC P51647; 009184;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase 1A1 (EC 1.2.1.3) (Aldehyde dehydrogenase,
DE cytosolic) (ALDH class 1) (Retinal dehydrogenase) (RALDH) (RALDH1).
GN ALDH1A1 OR ALDH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney;
RX MEDLINE=96125208; PubMed=8543180;
RA Bhat P.V., Labrecque J., Boutin J.-M., Lacroix A., Yoshida A.;
RT "Cloning of a cDNA encoding rat aldehyde dehydrogenase with high
RT activity for retinal oxidation.";
RL Gene 166:303-306 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=97212790; PubMed=9059608;
RA Kathmann E.C., Lipsky J.J.;
RT "A preliminary report on the cloning of a constitutively expressed rat
RT liver cytosolic ALDH cDNA by PCR.";
RL Adv. Exp. Med. Biol. 414:69-72 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Napoli J.L., Penzes P., Wang X., Sperkova Z.;
RT "Cloning of a rat cDNA encoding retinal dehydrogenase isozyme type
RT I and its expression in E. coli.";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-18; 79-90; 95-120; 204-224; 236-257 AND 393-437.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney;
RX MEDLINE=95134257; PubMed=7832787;
RA Labrecque J., Dumas F., Lacroix A., Bhat P.V.;
RT "A novel isoenzyme of aldehyde dehydrogenase specifically involved in

```

```

RT the biosynthesis of 9-cis and all-trans retinoic acid.";
RL Biochem. J. 305:681-684 (1995).
CC - FUNCTION: IS CAPABLE OF CONVERTING 9-CIS AND ALL-TRANS RETINAL TO
CC CORRESPONDING RETINOIC ACID WITH HIGH EFFICIENCY, 9-CIS RETINAL
CC BEING 2-FOLD MORE ACTIVE THAN ALL-TRANS RETINAL.
CC - CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC - ENZYME REGULATION: INHIBITED BY CHLORAL HYDRATE.
CC - PATHWAY: Ethanol utilization; second step.
CC - SUBUNIT: Homotetramer (By similarity).
CC - SUBCELLULAR LOCATION: Cytoplasmic.
CC - TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, LUNG, TESTIS,
CC INTESTINE, STOMACH, AND TRACHEA, BUT WEAKLY IN THE LIVER.
CC - SIMILARITY: Belongs to the aldehyde dehydrogenase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L42009; AAA96657.1; -.
DR EMBL; AF001896; AAC53304.1; -.
DR EMBL; AF001898; AAC53306.1; -.
DR EMBL; AF001897; AAC53305.1; -.
DR EMBL; U79118; AAB63423.1; -.
DR EIR; JC4524; JC4524.
DR PIR; JC5553; JC5553.
DR HSSP; P51977; 1BX5.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Acetylation.
FT INIT MET 0 0
FT MOD_RES 1 1 ACETYLATION (POTENTIAL).
FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 268 268 BY SIMILARITY.
FT ACT_SITE 302 302 BY SIMILARITY.
FT CONFLICT 99 99 R -> C (IN REF. 1).
FT CONFLICT 105 105 I -> M (IN REF. 4).
FT CONFLICT 169 169 N -> E (IN REF. 1).
SQ SEQUENCE 500 AA; 54327 MW; 93614C21A94B430D CRC64;

Query Match 1.7%; Score 7; DB 1; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 SPAQPAV 233
DB 2 SPAQPAV 8
|||||

RESULT 50
DHAI RAT
ID DHAI RAT STANDARD; PRT; 500 AA.
AC P13601;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase, cytosolic 1 (EC 1.2.1.3) (ALDH class 1)
DE (ALDH1) (ALDH-E1).
GN ALDH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89527272; PubMed=2753900;
RA Dunn T.J., Koleske A.J., Lindahl R., Pitot H.C.;
RT "Phenobarbital-inducible aldehyde dehydrogenase in the rat. cDNA

```



```

RT sequence and regulation of the mRNA by phenobarbital in responsive
RL rats."
CC J. Biol. Chem. 264:13057-13065 (1989).
CC -1- FUNCTION: BINDS FREE RETINAL AND CELLULAR RETINOL-BINDING PROTEIN-
CC BOUND RETINAL. CAN CONVERT/OXIDIZE RETINALDEHYDE TO RETINOIC ACID
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -1- PATHWAY: Ethanol utilization; second step.
CC -1- SUBUNIT: Homotetramer.
CC -1- INDUCTION: By phenobarbital.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M23995; AAA40718.1; -.
CC DR PIR; A32616; A32616.
CC DR HSSP; P51977; 1BX5.
CC DR InterPro; IPR002086; Aldehyde_dehydr.
CC DR Pfam; PF00171; aldehyd; 1.
CC DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
CC DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
CC KW Oxidoreductase; NAD.
CC FT INIT MET 0 BY SIMILARITY.
CC FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
CC FT ACT_SITE 268 268 BY SIMILARITY.
CC FT ACT_SITE 302 302 BY SIMILARITY.
CC SQ SEQUENCE 500 AA; 54428 MW; D6806A1AE29A0D56 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 227 SPAQPAV 233
Db 2 SPAQPAV 8
|||||

RESULT 51
ID IRX3 MOUSE STANDARD; PRT; 507 AA.
AC P81067.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Iroquois-class homeodomain protein IRX-3 (Iroquois homeobox protein
DE 3).
DN IRX3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98145944; PubMed=9486539;
RA Boase A., Zulch A., Becker M.B., Torres M., Gomez-Skarmeta J.-L.,
RA Modolell J., Gruss P.;
RA "Identification of the vertebrate Iroquois homeobox gene family with
RT overlapping expression during early development of the nervous
RT system.";
RL Mech. Dev. 69:169-181(1997).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE TALE/IRO HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y15001; CRA75233.1; -.
CC DR HSSP; P41778; 1DU6.
CC DR TRANSFAC; T02439; -.
CC DR MGD; MGI:1197522; Irx3.
CC DR InterPro; IPR001356; Homeobox.
CC DR InterPro; IPR003893; Iroquois_homeo.
CC DR Pfam; PF00046; homeobox; 1.
CC DR ProDom; PD000010; Homeobox; 1.
CC DR SMART; SM00389; HOX; 1.
CC DR SMART; SM00548; IRO; 1.
CC DR PROSITE; PS00027; HOMEBOX_1; 1.
CC DR PROSITE; PS00071; HOMEBOX_2; 1.
CC KW Homeobox; DNA-binding; Nuclear protein.
CC FT DNA_BIND 130 192 HOMEBOX (TALE-TYPE).
CC FT DOMAIN 24 29 POLY-GLY.
CC FT DOMAIN 64 73 POLY-ALA.
CC FT DOMAIN 228 235 POLY-GLU.
CC FT DOMAIN 250 253 POLY-ARG.
CC FT DOMAIN 315 319 POLY-PRO.
CC FT DOMAIN 391 399 POLY-ALA.
CC FT DOMAIN 451 458 POLY-ALA.
CC SQ SEQUENCE 507 AA; 52694 MW; C785AFEEA83E22FF CRC64;

Query Match 1.7%; Score 7; DB 1; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 PVARAPA 216
Db 319 PVARAPA 325
|||||

RESULT 52
ID NQOD PARDE STANDARD; PRT; 513 AA.
AC P29925.
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-quinone oxidoreductase chain 13 (EC 1.6.99.5) (NADH dehydrogenase
DE I. chain 13) (NDH-1, chain 13).
DN NQO13.
OS Paracoccus denitrificans.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Paracoccus.
OX NCBI_TaxID=266;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 13543;
RX MEDLINE=93136200; PubMed=8422400;
RA Xu X., Matsuno-Yagi A., Yagi T.;
RA "DNA sequencing of the seven remaining structural genes of the gene
RT cluster encoding the energy-transducing NADH-quinone oxidoreductase
RT of Paracoccus denitrificans.";
RL Biochemistry 32:968-981(1993).
CC -1- FUNCTION: NDH-1 shuttles electrons from NADH, via FMN and iron-
CC sulfur (Fe-S) centers, to quinones in the respiratory chain. The
CC immediate electron acceptor for the enzyme in this species is
CC believed to be ubiquinone. Couples the redox reaction to proton
CC translocation (for every two electrons transferred, four hydrogen
CC ions are translocated across the cytoplasmic membrane), and thus
CC conserves the redox energy in a proton gradient.
CC -1- CATALYTIC ACTIVITY: NADH + quinone = NAD(+) + quinol.
CC -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS. SUBUNITS NQO7-14
CC CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: TO POLYPEPTIDE 4 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC OF CHLOROPLASTS OR MITOCHONDRIA.

```



```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L02334; AAA25599.1; -
CC PIR; I45456; I45456.
CC InterPro; IPR003918; NADHub oxford4.
CC InterPro; IPR001750; Oxidored_g1.
CC Pfam; PF00361; oxidored_g1; 1.
CC PRINTS; PR01437; NUOXDRDFA54.
CC Oxidoreductase; NAD; Quinone; Ubiquinone; Transmembrane.
KW TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT TRANSMEM 211 231 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT TRANSMEM 277 297 POTENTIAL.
FT TRANSMEM 312 332 POTENTIAL.
FT TRANSMEM 340 360 POTENTIAL.
FT TRANSMEM 383 403 POTENTIAL.
FT TRANSMEM 418 438 POTENTIAL.
FT TRANSMEM 463 483 POTENTIAL.
SQ SEQUENCE 513 AA; 56417 MW; 8164DEACA20FE739 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 OPAAPVA 104
Db 501 OPAAPVA 507
|||||

RESULT 53
THS3_HALVO
ID THS3_HALVO STANDARD; PRT; 524 AA.
AC Q9HHA2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thermosome subunit 3 (Heat shock protein CCT3).
GN CCT3.
OS Halobacterium volcanii (Haloflex volcanii).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Haloflex.
OX NCBI_TaxID=2246;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DS70;
RA Kovacs E., Lund P.A.;
RT "Sequence of the cct3 gene, a third chaperonin homolog in the
RT halophilic archaeon Haloflex volcanii.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MOLECULAR CHAPERONE; BINDS UNFOLDED POLYPEPTIDES IN
CC VITRO, AND HAS A WEAK ATPASE ACTIVITY (BY SIMILARITY).
CC -1- SUBUNIT: FORMS AN OLIGOMERIC COMPLEX OF EIGHT-MEMBERED RINGS
CC (BY SIMILARITY).
CC -1- INDUCTION: By heat shock.
CC -1- SIMILARITY: Belongs to the TCP-1 chaperonin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC -----
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF298660; AAG17906.1; -
CC HSSP; P48424; I46D.
CC InterPro; IPR002194; Chaperonin_TCP-1.
CC InterPro; IPR001844; Chaperonin_Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60_TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00750; TCP1_1; 1.
CC PROSITE; PS00751; TCP1_2; 1.
CC PROSITE; PS00995; TCP1_3; 1.
KW Chaperone; ATP-binding; Heat shock.
SQ SEQUENCE 524 AA; 55261 MW; CBEL2E2EAF91181E9 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 524;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 GVPFGAG 272
Db 404 GVPFGAG 410
|||||

RESULT 54
PTB_MOUSE
ID PTB_MOUSE STANDARD; PRT; 527 AA.
AC P17225;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Polypyrimidine tract-binding protein 1 (PTB) (Heterogeneous nuclear
DE ribonucleoprotein I) (hnRNP I).
GN PTBP1 OR PTB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 312-341.
RX MEDLINE=92105132; PubMed=1722210;
RA Bothwell A.L.M., Ballard D.W., Philbrick W.M., Lindwall G.,
RA Maher S.E., Bridgett M.M., Jamison S.P., Garcia-Bianco M.A.;
RT "Murine polypyrimidine tract binding protein. Purification, cloning,
RT and mapping of the RNA binding domain.";
RL J. Biol. Chem. 266:24657-24663(1991).
CC -1- FUNCTION: PLAYS A ROLE IN PRE-MRNA SPLICING. BINDS TO THE
CC POLYPYRIMIDINE TRACT OF INTRONS. MAY PROMOTE THE BINDING OF U2
CC SNRNP TO PRE-MRNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE C-TERMINAL 195 AMINO ACIDS OF PTB ARE SUFFICIENT FOR
CC SPECIFIC RNA BINDING.
CC -1- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52101; CAA36321.1; -
CC MGD; MGI:97791; Ptbpl.
CC GO; GO:0005634; C:nucleus; IDA.
CC InterPro; IPR006536; hnRNP-L_PTB.
CC InterPro; IPR000504; RNA_rec_mot.
CC Pfam; PF00076; rrm; 4.
CC SMART; SM00360; RRM; 3.
CC TIGRFAMs; TIGR01649; hnRNP-L_PTB; 1.
CC PROSITE; PS50102; RRM; 4.

```

```

DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW Nuclear protein; RNA-binding; mRNA splicing; Repeat.
FT DOMAIN 58 142 RNA-BINDING (RRM) 1.
FT DOMAIN 183 259 RNA-BINDING (RRM) 2.
FT DOMAIN 335 386 RNA-BINDING (RRM) 3.
FT DOMAIN 450 525 RNA-BINDING (RRM) 4.
FT DOMAIN 315 321 POLY-ALA.
SQ SEQUENCE 527 AA; 56478 MW; F18FD376010D76A CRC64;

Query Match
Best Local Similarity 100.0%; DB 1; Length 527;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 QFQALLQ 290
Db 220 QFQALLQ 226
|||||

RESULT 55
PTB_HUMAN
AC P26599; STANDARD; PRT; 531 AA.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polypyrimidine tract-binding protein 1 (PTB) (Heterogeneous nuclear
DE ribonucleoprotein I) (hnRNP I) (57 kDa RNA-binding protein PTB-1).
GN PTB1 OR PTB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 123-130 AND 219-238.
RC TISSUE=Placenta;
RX MEDLINE=91293583; PubMed=1906035;
RA Gil A., Sharp P.A., Jamison S.P., Garcia-Blanco M.A.;
RT "Characterization of cDNAs encoding the polypyrimidine tract-binding
RT protein.";
RL Genes Dev. 5:1224-1236 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91293584; PubMed=1906036;
RA Patton J.G., Mayer S.A., Tempst P., Nadal-Ginard B.;
RT "Characterization and molecular cloning of polypyrimidine
RT tract-binding protein: a component of a complex necessary for
RT pre-mRNA splicing.";
RL Genes Dev. 5:1237-1251 (1991).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92350668; PubMed=1641332;
RA Ghetti A., Pinol-Roma S., Michael W., Morandi C., Dreyfuss G.;
RT "hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear
RT localization and association with hnRNAs.";
RL Nucleic Acids Res. 20:3671-3678 (1992).
RN [4]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attix C., Andreise T., Trankheim M., Amico-Keller G., Cosfield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
RT serine protease gene cluster.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 123-129; 399-405; 411-426 AND 429-437.
RC TISSUE=Keratinocytes;
RX MEDLINE=93162043; PubMed=1286667;

```

```

RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandeckerkhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RT protein database of normal human epidermal keratinocytes.";
EL Electrophoresis 13:960-969 (1992).
CC -1- FUNCTION: PLAYS A ROLE IN PRE-MRNA SPLICING. BINDS TO THE
CC POLYPYRIMIDINE TRACT OF INTRONS. MAY PROMOTE THE BINDING OF U2
CC SNRNP TO PRE-MRNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P26599-1; Sequence=Displayed;
CC Name=2; Synonyms=PTB2;
CC IsoId=P26599-2; Sequence=VSP_005802;
CC -1- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
DR EMBL; X62006; CAA43973.1; -
DR EMBL; X65371; CAA46443.1; -
DR EMBL; X65372; CAA46444.1; -
DR EMBL; X60648; CAA43056.1; -
DR EMBL; X66975; CAA47386.1; -
DR EMBL; AC006273; AAC99798.1; -
DR PIR; S23016; S23016.
DR PDB; 1QM9; 03-JUL-00.
DR Aarnus/Ghent-2DFAGE; 1505; NEPHGE.
DR Genew; HGNC:9583; PTBP1.
DR GK; P26599; -
DR MIM; 600693; -
DR GO; GO:0005730; C:nucleolus; TAS.
DR GO; GO:0005654; C:nucleoplasm; TAS.
DR GO; GO:0008436; F:heterogeneous nuclear ribonucleoprotein; TAS.
DR GO; GO:0008187; F:poly-pyrimidine tract binding activity; TAS.
DR GO; GO:0006371; P:mRNA splicing; TAS.
DR InterPro; IPR006536; HnRNP-L_PTB.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 4.
DR SMART; SM00360; RRM; 4.
DR TIGRFAMs; TIGR01649; hnRNP-L_PTB; 1.
DR PROSITE; PS0102; RRM; 4.
DR PROSITE; PS0030; RRM_RNP_1; FALSE_NEG.
KW Nuclear protein; RNA-binding; mRNA splicing; Repeat;
KW Alternative splicing; 3D-structure.
FT DOMAIN 59 143 RNA-BINDING (RRM) 1.
FT DOMAIN 184 260 RNA-BINDING (RRM) 2.
FT DOMAIN 337 411 RNA-BINDING (RRM) 3.
FT DOMAIN 454 529 RNA-BINDING (RRM) 4.
FT DOMAIN 316 323 POLY-ALA.
FT VARSPLIC 297 297 F -> FASPYAGAGPPTFAIPQAA (in isoform 2).
SQ SEQUENCE 531 AA; 57221 MW; BE12D5EA21537BED CRC64;
/FTID=VSP_005802.

Query Match
Best Local Similarity 100.0%; DB 1; Length 531;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 QFQALLQ 290
Db 221 QFQALLQ 227
|||||

RESULT 56
GAG_BAEVM
ID _GAG_BAEVM STANDARD; PRT; 537 AA.

```

P03341; P10268;  
 21-JUL-1986 (Rel. 01, Created)  
 01-MAR-1989 (Rel. 10, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 GAG polyprotein [contains: inner coat protein P12; Core protein P15;  
 Core shell protein P30; Nucleoprotein P10].  
 GAG.  
 GN Baboon endogenous virus (strain M7).  
 OC Viruses; Retrovirdae; Retroviridae; Mammalian type C retroviruses.  
 OX NCBI\_TaxID=11764;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kato S., Matsuo K., Nishimura N., Takahashi N., Takano T.;  
 RT "The entire nucleotide sequence of baboon endogenous virus DNA: a  
 RT chimeric genome structure of murine type C and simian type D  
 RT retroviruses."; J. Virol. 47:137-145(1983).  
 RL Jpn. J. Genet. 62:127-137(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=83241915; PubMed=6408267;  
 RX Tamura T.;  
 RT "Provirus of M7 baboon endogenous virus: nucleotide sequence of the  
 RT gag-pol region."; J. Virol. 47:137-145(1983).  
 RL J. Virol. 47:137-145(1983).  
 CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
 CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL  
 CC POLYPROTEIN.  
 CC -!- SIMILARITY: Contains 1 CCHC-type zinc finger.  
 CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; D10032; BAA00923.1; -;  
 DR EMBL; X05470; CAA29027.1; -;  
 DR EMBL; J02034; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; J02060; FOMVM7.  
 DR InterPro; IPR000840; Gag\_MA.  
 DR InterPro; IPR003036; Gag\_P30.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF01140; Gag\_MA; 1.  
 DR Pfam; PF02093; Gag\_P30; 1.  
 DR Pfam; PF00098; Znf\_CCHC; 1.  
 DR SMART; SM00343; Znf\_CCHC; 1.  
 DR PROSITE; PS0158; ZF\_CCHC; 1.  
 DR Core protein; Coat protein; Nucleoprotein; Polyprotein; Myristate;  
 KW Zinc-finger.  
 FT CHAIN 2 112 INNER COAT PROTEIN P12.  
 FT CHAIN 113 226 CORE PROTEIN P15.  
 FT CHAIN 227 477 CORE SHELL PROTEIN P30.  
 FT CHAIN 478 537 NUCLEOPROTEIN P10.  
 FT ZN FING 501 518 CCHC-TYPE.  
 FT LIPID 2 2 MYRISTATE.  
 SQ SEQUENCE 537 AA; 60623 MW; 5DEEE4437CFCFB79 CRC64;  
 Query Match 1.7%; Score 7; DB 1; Length 537;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 355 TPEEREA 361  
 DB 446 TPEEREA 452  
 RESULT 57  
 SYK\_RHILO  
 ID SYK\_RHILO STANDARD; PRT; 548 AA.  
 AC Q98EC8;

28-FEB-2003 (Rel. 41, Created)  
 28-FEB-2003 (Rel. 41, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 Lysyl-tRNA synthetase [EC 6.1.1.6] (Lysine--tRNA ligase) (LysRS).  
 GN LYS OR MLJ5634.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti"; DNA Res. 7:331-338(2000).  
 RL DNA Res. 7:331-338(2000).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate  
 CC + L-lysyl-tRNA(Lys).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AP003007; BAB52044.1; -;  
 DR HAMAP; MF\_00177; -; 1.  
 DR InterPro; IPR002904; Lys tRNA-synt\_1c.  
 DR InterPro; IPR001412; tRNA-synt\_1.  
 DR Pfam; PF01921; tRNA-synt\_1f; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.  
 FT SITE 52 60 "HIGH" REGION.  
 FT SITE 300 304 "KMSK" REGION.  
 FT BINDING 303 303 ATP (BY SIMILARITY).  
 SQ SEQUENCE 548 AA; 61014 MW; BSB44B8DF3D89C29 CRC64;  
 Query Match 1.7%; Score 7; DB 1; Length 548;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 331 SPEGGPG 337  
 DB 495 SPEGGPG 501  
 RESULT 58  
 SYK\_BRUME  
 ID SYK\_BRUME STANDARD; PRT; 551 AA.  
 AC Q8YCM8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Lysyl-tRNA synthetase [EC 6.1.1.6] (Lysine--tRNA ligase) (LysRS).  
 GN LYS OR BME110500.  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OX NCBI\_TaxID=29459;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16M / ATCC 23456 / Biotype 1;  
 RX MEDLINE=20020109; PubMed=1175668;

```

RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Muder C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Hagelkorn R., Kyrides N., Overbeek R.,
RA "The genome sequence of the facultative intracellular pathogen
RA Brucella melitensis.";
RA Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -1- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
CC + L-lysyl-tRNA(Lys).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE009686; AA153742.1; -.
CC PIR; AC3572; AC3572.
CC HANAP; MF_00177; -.
CC InterPro; IPR002904; Lys-trna-synt_1c.
CC InterPro; IPR001412; trna-synt_1.
CC Pfam; PF01921; trna-synt_1f; 1.
CC PROSITE; PS00178; AA TRNA_LIGASE I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 54 62 "HIGH" REGION.
CC SITE 303 307 "KMSKS" REGION.
CC BINDING 306 306 ATP (BY SIMILARITY).
CC SEQUENCE 551 AA; 61699 MW; FC55DB8B469E1ACO CRC64;
CC
CC Query Match 1.7%; Score 7; DB 1; Length 551;
CC Best Local Similarity 100.0%; Pred. No. 1.2e+02;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 331 SPEGGPG 337
CC DB 498 SPEGGPG 504
CC
CC RESULT 59
CC ID SYK BRUSU STANDARD; PRT; 551 AA.
CC AC PS9225;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DE Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) (LysRS).
CC GN LYSS OR BRA0790.
CC OS Brucella suis.
CC OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
CC OC Brucellaceae; Brucella.
CC OX NCBI_TaxID=29461;
CC [1]
CC SEQUENCE FROM N.A.
CC RC STRAIN=1330 / Biovar 1;
CC RX MEDLINE=22247741; PubMed=12271122;
CC RA Paulsen I.T., Seshadri R., Nelson K.E., Brinkac L.M., Beanan M.J.,
CC Read T.D., Dodson R.J., Unayam L., Kolonay J.F., Madupu R.,
CC Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
CC Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.E.,
CC Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
CC Hoover D.L., Lindler L.E., Helling S.M., Boyle S.M., Fraser C.M.;
CC "The Brucella suis genome reveals fundamental similarities between
CC animal and plant pathogens and symbionts.";
CC Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
CC -1- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
CC + L-lysyl-tRNA(Lys).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

```

```

CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014573; AA33969.1; -.
CC TIGR; BRA0790; -.
CC HANAP; MF_00177; -.
CC PROSITE; PS00178; AA TRNA_LIGASE I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SITE 54 62 "HIGH" REGION.
CC SITE 303 307 "KMSKS" REGION.
CC BINDING 306 306 ATP (BY SIMILARITY).
CC SEQUENCE 551 AA; 61668 MW; 5E9B3CAB71A0EB80 CRC64;
CC
CC Query Match 1.7%; Score 7; DB 1; Length 551;
CC Best Local Similarity 100.0%; Pred. No. 1.2e+02;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 331 SPEGGPG 337
CC DB 498 SPEGGPG 504
CC
CC RESULT 60
CC ID PTB RAT STANDARD; PRT; 555 AA.
CC AC Q00438; Q63568;
CC DT 01-APR-1993 (Rel. 25, Created)
CC DT 01-APR-1993 (Rel. 25, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Polypyrimidine tract-binding protein 1 (PTB) (Heterogeneous nuclear
CC ribonucleoprotein 1) (hnRNP I) (Pyrimidine-binding protein) (PYBP).
CC GN PTBP1 OR PTB OR PYBP OR TEFII.
CC OS Rattus norvegicus (Rat).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC OX NCBI_TaxID=10116;
CC [1]
CC SEQUENCE FROM N.A., AND SEQUENCE OF 185-205; 348-365 AND 517-549.
CC TISSUE=Hepatoma;
CC RX MEDLINE=92020211; PubMed=1681508;
CC RA Brunel F., Alzari P.M., Ferrara P., Zakari M.M.;
CC "Cloning and sequencing of PYBP, a pyrimidine-rich specific single
CC strand DNA-binding protein.";
CC Nucleic Acids Res. 19:5237-5245(1991).
CC [2]
CC SEQUENCE FROM N.A.
CC RA Sengupta P.;
CC RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS A ROLE IN PRE-MRNA SPLICING. BINDS TO THE
CC POLYPYRIMIDINE TRACT OF INTRONS. MAY PROMOTE THE BINDING OF U2
CC SNRNP TO PRE-MRNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Names=PYBP2;
CC IsoIdc=Q00438-1; Sequence=Displayed;
CC Names=PYBP1;
CC IsoIdc=Q00438-2; Sequence=VSP_005803;
CC -1- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X60789; CAA43202.1; -;  
 DR EMBL; X60790; CAA43203.1; -;  
 DR EMBL; X74565; CAA52653.1; -;  
 DR PIR; S15522; S15552.  
 DR PIR; S36629; S36629.  
 DR InterPro; IPR000536; HnRNP-L\_PTB.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 4.  
 DR SMART; SM00360; RRM; 4.  
 DR TIGRfams; TIGR01649; hnRNP-L\_PTB; 1.  
 DR PROSITE; PS0102; RRM; 4.  
 DR PROSITE; PS00030; RRM\_RNP\_1; FALSE NEG.  
 KW Nuclear protein; RNA-binding; mRNA splicing; Repeat;  
 KW Alternative splicing.  
 FT DOMAIN 58 142 RNA-BINDING (RRM) 1.  
 FT DOMAIN 183 259 RNA-BINDING (RRM) 2.  
 FT DOMAIN 361 435 RNA-BINDING (RRM) 3.  
 FT DOMAIN 478 553 RNA-BINDING (RRM) 4.  
 FT DOMAIN 340 347 POLY-ALA.  
 FT VARSPIC 297 321 Missing (in isoform PYBP1).  
 FT CONFLICT 57 57 S -> T (IN REF. 2).  
 FT CONFLICT 164 164 A -> R (IN REF. 2).  
 FT CONFLICT 311 320 VPSHLCHPSR -> GFPTFAIPAQA (IN REF. 2).  
 FT CONFLICT 542 549 ENHHLRVS -> LGDNHKKR (IN REF. 1; AA  
 SEQUENCE).  
 SQ SEQUENCE 555 AA; 59353 MW; E73F22B54467117F CRC64;

Query Match 1.7%; Score 7; DB 1; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 284 QFOALLQ 290  
 DB 220 QFOALLQ 226

RESULT 61  
 PTB\_PIG STANDARD; PRT; 557 AA.  
 AC Q29099;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Polypyrimidine tract-binding protein 1 (PTB) (Heterogeneous nuclear  
 ribonucleoprotein I) (hnRNP I).  
 GN PTBP1 OR PTB.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=96249475; PubMed=8654585;  
 RA Niepmann M.;  
 RT "Portine polypyrimidine tract-binding protein stimulates translation  
 initiation at the internal ribosome entry site of foot-and-mouth-  
 disease virus.";  
 RL FEBS Lett. 388:39-42(1996).  
 CC -1- FUNCTION: PLAYS A ROLE IN PRE-MRNA SPLICING. BINDS TO THE  
 CC POLYPYRIMIDINE TRACT OF INTRONS. MAY PROMOTE THE BINDING OF U2  
 CC SNRNP TO PRE-MRNA.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X93009; CAA63597.1; -;  
 DR PIR; S68857; S68857.  
 DR InterPro; IPR006536; HnRNP-L\_PTB.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 4.  
 DR SMART; SM00360; RRM; 4.  
 DR TIGRfams; TIGR01649; hnRNP-L\_PTB; 1.  
 DR PROSITE; PS0102; RRM; 4.  
 DR PROSITE; PS00030; RRM\_RNP\_1; FALSE NEG.  
 KW Nuclear protein; RNA-binding; mRNA splicing; Repeat.  
 FT DOMAIN 59 143 RNA-BINDING (RRM) 1.  
 FT DOMAIN 184 260 RNA-BINDING (RRM) 2.  
 FT DOMAIN 363 437 RNA-BINDING (RRM) 3.  
 FT DOMAIN 480 555 RNA-BINDING (RRM) 4.  
 FT DOMAIN 342 349 POLY-ALA.  
 SQ SEQUENCE 557 AA; 59855 MW; E323F8FBD6DAFF66 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 557;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 284 QFOALLQ 290  
 DB 221 QFOALLQ 227

RESULT 62  
 YHJU\_ECOLI STANDARD; PRT; 559 AA.  
 ID YHJU\_ECOLI  
 AC P37659;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yhjU.  
 GN YHJU OR B3538.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 RT region from 76.0 to 81.5 minutes.";  
 RL Nucleic Acids Res. 22:2576-2586(1994).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U00039; AAB18516.1; -;  
 DR EMBL; AE000431; AAC76563.1; -;  
 DR PIR; S47760; S47760.  
 DR Ecogene; Egi2265; yhjU.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 559 AA; 62032 MW; 7528D897B0E3E4E9 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 559;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 135 AATVAAT 141

```
Db 179 AATVAAT 185
|||||
RESULT 63
Y745 ARATH STANDARD; PRT; 597 AA.
AC F59278; Q9C812;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein At1g51745.
GN At1g51745 OR F19C24.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egtu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizler L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.B., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uteckebat T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820 (2000).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT SSP consortium (Salk/Stanford/RGEC).";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 PWWP domain.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.iesb-sib.ch/announce/
CC or send an email to license@iesb-sib.ch).
CC
CC EMBL; AC025294; AAG50888.1; AUT_SEQ.
DR EMBL; BT002989; AAO22798.1; -.
DR PROSITE; PS50812; PWWP; 1.
KW Hypothetical protein.
FT DOMAIN 16 78 PWWP.
FT CONFLICT 268 268 M -> R (IN REF. 1).
SQ SEQUENCE 597 AA; 65044 MW; E59141F5B9DFD79F CRC64;

Query Match 1.7%; Score 7; DB 1; Length 597;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 76 KAKASSS 82

Db 179 AATVAAT 185
|||||
RESULT 64
SYR_PYRO STANDARD; PRT; 629 AA.
AC O59147;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine-tRNA ligase) (ArgRS).
GN ARGS OR PH1478.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76 (1998).
CC -1- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.iesb-sib.ch/announce/
CC or send an email to license@iesb-sib.ch).
CC
CC EMBL; AP000006; BAA30585.1; -.
DR PIR; A71023; A71023.
DR HAMAP; MF 00123; -.
DR InterPro; IPR001278; Arg_tRNA-synt_1c.
DR InterPro; IPR005148; N.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF03485; N-Arg; 1.
DR Pfam; PF00750; tRNA-synt 1d; 1.
DR PRINTS; PR01038; TRNASYNTHARG.
DR TIGRFAMs; TIGR00456; argS; 1.
DR PROSITE; PS00178; AA-TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
FT Complete proteome.
FT SITE 128 138 "HIGH" REGION.
SQ SEQUENCE 629 AA; 72339 MW; 34CE2D841DBAD48A CRC64;

Query Match 1.7%; Score 7; DB 1; Length 629;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 LVIMLSK 76
|||||
Db 549 LVIMLSK 555

RESULT 65
COG6_DROME STANDARD; PRT; 630 AA.
ID COG6_DROME
AC Q9V564; Q95RW7;
DT 28-FEB-2003 (Rel. 41, Created)
```

28-FEB-2003 (Rel. 41, Last sequence update)  
15-SEP-2003 (Rel. 42, Last annotation update)  
Putative conserved oligomeric Golgi complex component 6.  
C91968.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.P.,  
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers J.H., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ithegam C.,  
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP REVISIONS, AND ALTERNATIVE SPLICING.  
RC STRAIN=Berkley;  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Battencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley; TISSUE=Embryo;  
RX MEDLINE=22426066; PubMed=12537569;  
RA George R.A., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA Stapleton M., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,  
RA Rubin G.M., Celniker S.E.;  
RT "A Drosophila full-length cDNA resource.";

Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
-!- FUNCTION: Required for normal Golgi function (By similarity).  
-!- SUBUNIT: Component of the conserved oligomeric Golgi complex which  
is composed of eight different subunits and is required for normal  
Golgi morphology and localization (By similarity).  
-!- SUBCELLULAR LOCATION: Golgi (By similarity).  
-!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Name=Long;  
IsoId=Q9V564-1; Sequence=Displayed;  
Name=Short;  
IsoId=Q9V564-2; Sequence=VSP\_001133, VSP\_001134;  
-!- SIMILARITY: BELONGS TO THE COG6 FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AB003834; AAF58956.1; -;  
EMBL; AB003834; AAF58956.1; -;  
EMBL; AY061082; AAL28630.1; -;  
FlyBase; FBgn0033401; CG1968.  
Hypothetical protein; Transport; Protein transport; Golgi stack;  
Membrane; Alternative splicing.  
VARSPPLIC 315 333 DISOLQNALGYIADGVCH -> GKIRKCIHYGTIVLPVL  
Q (in isoform Short).  
/FTid=VSP\_001133.  
FT VARSPLIC 334 630 Missing (in isoform Short).  
/FTid=VSP\_001134.  
FT SEQUENCE 630 AA; 71205 MW; 934113C7B266ABFD CRC64;  
Query Match 1.7%; Score 7; DB 1; Length 630;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 333 EGGPGGN 339  
DB 265 EGGPGGN 271  
-----  
RESULT 66  
PRIM\_UREPA STANDARD; PRT; 641 AA.  
ID AC Q9PPZ6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA primase (EC 2.7.7.-).  
GN DNAG OR UUA494.  
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.  
OX NCBI\_TaxID=134821;  
RN [1]\_TaxID=134821;  
RP SEQUENCE FROM N.A.  
RC STRAIN=Serovar 3;  
RX MEDLINE=20500219; PubMed=11048724;  
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
RA Cassell G.H.;  
RT "The complete sequence of the mucosal pathogen Ureaplasma  
urealyticum.";  
RL Nature 407:757-762(2000).  
CC -!- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL  
RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT  
REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.  
CC -!- COPACITOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).  
CC -!- SUBUNIT: Monomer (By similarity).  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration -  
between the Swiss Institute of Bioinformatics and the EMBL outstation -



```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE002147; AAF30906.1; -.
DR HSP; Q9X4D0; IDOQ.
DR InterPro; IPR006295; DNA_primase.
DR InterPro; IPR006171; Toprim_dom.
DR InterPro; IPR006647; Toprim_dom.
DR InterPro; IPR006154; Toprim_sub.
DR InterPro; IPR002694; Znf_CHC2.
DR Pfam; PF01751; Toprim; 1.
DR Pfam; PF01807; zf-CHC2; 1.
DR ProDom; PD002276; Toprim_primease; 1.
DR ProDom; PD002988; Znf_CHC2; 1.
DR SMART; SM00493; TOPRIM; 1.
DR TIGR; TIGR01391; dnaG; 1.
DR Transferase; DNA replication; DNA-directed RNA polymerase; Primosome;
KW Zinc-finger; Zinc; Metal-binding; Complete proteome.
FT ZN FING 39 63
FT CHC2-TYPE (BY SIMILARITY).
SQ SEQUENCE 641 AA; 75390 MW; A6920168DB58EF47 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 641;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ENSFLV1 72
| | | | |
DB 191 ENSFLV1 197

RESULT 67
VP40_HSVB
ID VP40_HSVB STANDARD; PRT; 646 AA.
AC P28936; Q69263;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Capsid protein P40 (Viron structural gene 35 protein) [Contains:
DE Capsid protein VP24 (Assemblin) (Protease) (EC 3.4.21.97); Capsid
DE protein VP22A)].
GN 35.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92295566; PubMed=1318506;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -!- FUNCTION: VP22A IS A COMPONENT OF THE CAPSID CORE INVOLVED IN
CC PROCESSING AND PACKAGING OF PROGENY DNA. VP24 IS A PROTEASE
CC WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF AND VP22A AT THE C-
CC TERMINUS.
CC -!- CATALYTIC ACTIVITY: Cleaves -Ala-|-Ser- and -Ala-|-Ala- bonds in
CC the scaffold protein.
CC -!- SIMILARITY: BELONGS TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M86664; AAB02470.1; -.
DR EMBL; M86664; AAB02471.1; -.

```

```

DR PIR; I36798; WZBEC8.
DR HSP; P16753; IWPO.
DR MEROPS; S21.001; -.
DR InterPro; IPR001847; Assemblin.
DR Pfam; PF00716; Peptidase_S21; 1.
DR PRINTS; PR00236; HSCVAPSIDP40.
KW Coat protein; Hydrolase; Serine protease.
FT CHAIN 1 646
FT CHAIN 318 646
FT CHAIN 1 242
FT CHAIN 243 7622
FT PROPEP 7623 646
FT SITE 242 243
FT SITE 622 623
FT ACT_SITE 55 55
FT ACT_SITE 123 123
FT ACT_SITE 142 142
FT ACT_SITE 646 646
SQ SEQUENCE 646 AA; 69579 MW; 1C1C28C848116276 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 646;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 QPQAAPA 135
| | | | |
DB 513 QPQAAPA 519

RESULT 68
PKNI_COREF
ID PKNI_COREF STANDARD; PRT; 660 AA.
AC Q8FUI5;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable serine/threonine protein kinase CE0033 (EC 2.7.1.37).
GN CE0033.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kwarabavasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -!- SIMILARITY: Contains 3 PASTA domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AP005214; BAC16843.1; -.
DR InterPro; IPR005543; PASTA.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF03793; PASTA; 3.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00740; PASTA; 3.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

```



DR PROSITE; PS00101; PROTEIN KINASE DOM; 1.  
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;  
KW ATP-binding; Repeat; Complete proteome.  
FT DOMAIN 9 278 PROTEIN KINASE.  
FT DOMAIN 377 443 PASTA 1.  
FT DOMAIN 444 512 PASTA 2.  
FT DOMAIN 513 577 PASTA 3.  
FT NP BIND 15 23 ATP (BY SIMILARITY).  
FT BINDING 38 38 ATP (BY SIMILARITY).  
FT ACT\_SITE 136 136 BY SIMILARITY.  
SQ SEQUENCE 660 AA; 69646 MW; D33D797EB02D44B1 CRC64;  
  
Query Match 1.7%; Score 7; DB 1; Length 660;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 134 PAATAVA 140  
Db 318 PAATAVA 324  
|||||  
318 PAATAVA 324  
  
RESULT 69  
SCRT DROME STANDARD; PRT; 666 AA.  
AC P45843;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Scarlet protein.  
GN ST.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Canton-S.  
RA Garcia R.L., Perkins H.D., Howells A.J.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 200-306 FROM N.A.  
RX MEDLINE-89339145; PubMed-2503416;  
RA Tearle R.G., Belote J.M., McKeown M., Baker B.S., Howells A.J.;  
RT "Cloning and characterization of the scarlet gene of Drosophila melanogaster."  
RL Genetics 122:595-606(1989).  
CC -1- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS RESPONSIBLE FOR EYE COLOR. SCARLET AND WHITE DIMERIZE FOR THE TRANSPORT OF TRYPTOPHAN.  
CC -1- SUBUNIT: HETERODIMER OF SCARLET WITH WHITE.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY, MDR SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U39739; AAA82056.1; -  
DR EMBL; X76201; CAA53794.1; -  
DR FLYBase; FBgn0003515; st.  
DR GO; GO:0006727; P:omochrome biosynthesis; IMP.  
DR InterPro; IPR003593; AAA\_Alpase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR005284; Pigment\_permease.  
DR Pfam; PF000005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.

DR SMART; SM00382; AAA; 1.  
DR TIGRFAMs; TIGR00955; 3a01204; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
KW Pigment; ATP-binding; Transmembrane; Transport.  
FT NP BIND 108 115 ATP (POTENTIAL).  
FT TRANSMEM 418 438 POTENTIAL.  
FT TRANSMEM 445 465 POTENTIAL.  
FT TRANSMEM 491 511 POTENTIAL.  
FT TRANSMEM 519 539 POTENTIAL.  
FT TRANSMEM 552 572 POTENTIAL.  
FT TRANSMEM 577 597 POTENTIAL.  
FT TRANSMEM 640 660 POTENTIAL.  
SQ SEQUENCE 666 AA; 74506 MW; 6796ED4084B59CE4 CRC64;  
  
Query Match 1.7%; Score 7; DB 1; Length 666;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 251 LNLFPQG 257  
Db 462 LNLFPQG 468  
|||||  
462 LNLFPQG 468  
  
RESULT 70  
TUP1\_KLULA STANDARD; PRT; 682 AA.  
ID TUP1\_KLULA  
AC P56094;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Transcriptional repressor TUP1.  
GN TUP1.  
OS Kluyveromyces fragilis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
OX NCBI\_TaxID=28985;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Braun B.R., Johnson A.D.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE-20036803; PubMed-10567571;  
RA Mukai Y., Matsuo E., Roth S.Y., Harashima S.;  
RT "Conservation of histone binding and transcriptional repressor functions in a Schizosaccharomyces pombe Tup1 homolog."  
RL Mol. Cell. Biol. 19:8461-8468(1999).  
CC -1- FUNCTION: REPRESSOR TRANSCRIPTION BY RNA POLYMERASE II.  
CC -1- SIMILARITY: Contains 7 WD repeats.  
CC -1- SIMILARITY: BELONGS TO THE TUP1 FAMILY OF WD-REPEAT PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF005740; AAB63194.1; -  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 7.  
DR ProDom; PD000018; WD40; 4.  
DR SMART; SM00320; WD40; 7.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 5.  
DR PROSITE; PS00082; WD\_REPEATS\_2; 5.  
DR PROSITE; PS02944; WD\_REPEATS\_REGION; 1.  
KW Transcription regulation; Repressor; Repeat; WD repeat.  
FT REPEAT 319 359 WD 1.  
FT REPEAT 404 443 WD 2.  
FT REPEAT 446 485 WD 3.  
FT REPEAT 487 527 WD 4.

FT REPEAT 537 576 WD 5.  
 FT REPEAT 586 625 WD 6.  
 FT REPEAT 628 673 WD 7.  
 FT DOMAIN 172 180 POLY-GLN.  
 FT DOMAIN 184 194 POLY-GLN.  
 SQ SEQUENCE 682 AA; 74054 MW; E17C24A36232445E CRC64;

Query Match 1.7%; Score 7; DB 1; Length 682;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 ASTATTA 90  
 |||||  
 Db 233 ASTATTA 239

RESULT 71  
 EFG BRUME  
 ID EFG BRUME STANDARD; PRT; 694 AA.  
 AC Q8YHP3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Elongation factor G (EF-G).  
 GN FUSA OR EME10754.  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29459;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16M / ATCC 23456 / Biotype 1;  
 RX MEDLINE=20020109; PubMed=11756688;  
 RA DelVecchio V.G., Kapatal V., Redkar R.J., Patra G., Muijer C., Los T.,  
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,  
 RA Jablonski L., Larsen P.H., D'Souza M., Bernal A., Mazur M., Goltzman E.,  
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
 RA Haeckl Korn R., Kyrides N., Overbeek R.;  
 RT "The genome sequence of the facultative intracellular pathogen  
 Brucella melitensis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
 CC -1- FUNCTION: This protein promotes the GTP-dependent translocation of  
 the nascent protein chain from the A-site to the P-site of the  
 ribosome.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-G/EF-2 SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 CC EMBL: AE009518; AAL51935.1; --  
 CC PIR: AD3346; AD3346.  
 CC HAWAP: MF\_00054; -- 1.  
 CC InterPro: IPR004540; EF-G.  
 CC InterPro: IPR000795; EF GTPbind.  
 CC InterPro: IPR000640; EFG C.  
 CC InterPro: IPR00517; EFG-IV.  
 CC InterPro: IPR004161; EFTU D2.  
 CC InterPro: IPR005225; Small GTP.  
 CC Pfam: PF00679; EFG C; 1.  
 CC Pfam: PF03764; EFG IV; 1.  
 CC Pfam: PF00009; GTP\_EFTU; 1.  
 CC Pfam: PF03144; GTP\_EFTU\_D2; 1.  
 CC TIGRFAMs: TIGR00484; EF-G; 1.  
 CC TIGRFAMs: TIGR00231; small GTP; 1.  
 CC PROSITE: PS00301; EFACITOR GTP; 1.  
 KW Elongation factor; Protein biosynthesis; GTP-binding;

KW Complete proteome.  
 FT NP\_BIND 17 24 GTP (BY SIMILARITY).  
 FT NP\_BIND 86 90 GTP (BY SIMILARITY).  
 FT NP\_BIND 140 143 GTP (BY SIMILARITY).  
 SQ SEQUENCE 694 AA; 76295 MW; ED1303A991DC34E2 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 694;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 AVALPVQ 238  
 |||||  
 Db 162 AVALPVQ 168

RESULT 72  
 STE11 YEAST  
 ID STE11 YEAST STANDARD; PRT; 738 AA.  
 AC P23561;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Serine/threonine-protein kinase STE11 (EC 2.7.1.37).  
 GN STE11 OR YLR362W OR YLR039.10.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE OF 22-738 FROM N.A.  
 RX MEDLINE=91115076; PubMed=2276621;  
 RA Rhodes N., Connell L., Errede B.;  
 RT "STE11 is a protein kinase required for cell-type-specific  
 transcription and signal transduction in yeast.";  
 RL Genes Dev. 4:1862-1874(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97313267; PubMed=9169871;  
 RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Anseorge W.,  
 RA Benes V., Bruckner M., Delli H., Dubois E., Duesterhoft A.,  
 RA Etian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,  
 RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,  
 RA Louis E.J., Messenguy P., Mewes H.-W., Miesing T., Moestl D.,  
 RA Mueller-Ruer S., Nentwich U., Obermaier B., Piravandi E., Fohl T.M.,  
 RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,  
 RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,  
 RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,  
 RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,  
 RA Wedler H., Zimmermann P.K., Zollner A., Hani J., Hoheisel J.D.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";  
 RL Nature 387:87-90(1997).  
 RN [3]  
 RP POSSIBLE FUNCTION.  
 RX MEDLINE=92313935; PubMed=1628833;  
 RA Cairns B.R., Ramer S.W., Kornberg K.D.;  
 RT "Order of action of components in the yeast pheromone response  
 pathway revealed with a dominant allele of the STE11 kinase and the  
 multiple phosphorylation of the STE7 kinase.";  
 RL Genes Dev. 6:1305-1318(1992).  
 CC -1- FUNCTION: SERINE/THREONINE PROTEIN KINASE REQUIRED FOR CELL-TYPE-  
 SPECIFIC TRANSCRIPTION AND SIGNAL TRANSDUCTION IN YEAST. IT  
 IS THOUGHT THAT IT PHOSPHORYLATES THE STE7 PROTEIN KINASE WHICH  
 ITSELF, PHOSPHORYLATES THE FUS3 AND OR KSS1 KINASES.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -1- PATHWAY: RESPONSE TO PHEROMONE-INDUCED SIGNAL.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC MAP KINASE KINASE SUBFAMILY.  
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; X53431; CAA37522.1; -;  
DR EMBL; U19103; AAB67571.1; -;  
DR PIR; S51380; S51380.  
DR SGD; S0004354; STE11.  
DR GO; GO:0005737; Cytoplasm; IDA.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR001660; SAM.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR Pfam; PF00069; kinase; 1.  
DR Pfam; PF00536; SAM; 1.  
DR PRINTS; PR00109; TYRKINASE.  
DR PRODOM; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR SMART; SM00454; SAM; 1.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS00105; SAM\_DOMAIN; 1.  
DR Transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Phosphonome response.  
FT DOMAIN 41 105 SAM.  
FT DOMAIN 436 733 PROTEIN KINASE.  
FT NP\_BIND 442 450 ATP (BY SIMILARITY).  
FT BINDING 465 465 ATP (BY SIMILARITY).  
FT ACT\_SITE 600 600 BY SIMILARITY.  
SQ SEQUENCE 738 AA; 83217 MW; A50D69800B346A41 CRC64;

Query Match 1.78; Score 7; DB 1; Length 738;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 TATTAKA 92  
|||||||  
Db 653 TATTAKA 659

## RESULT 73

ID\_CHICK  
EL ELS CHICK STANDARD; PRT; 750 AA.  
AC P07916;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Elastin precursor (Tropoelastin) (Fragment).  
GN ELN.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87242320; PubMed=3593675;  
RA Bressan G.M., Argos P., Stanley K.K.;  
RT "Repeating structure of chick tropoelastin revealed by complementary  
RT DNA cloning";  
RL Biochemistry 26:1497-1503(1987).  
RN [2]  
RP SEQUENCE OF 85-750 FROM N.A. (ISOFORM 2).  
RX MEDLINE=88309083; PubMed=2841924;  
RA Baule V.J., Foster J.A.;  
RT "Multiple chick tropoelastin mRNAs";  
RL Biochem. Biophys. Res. Commun. 154:1054-1060(1988).  
RN [3]  
RP SEQUENCE OF 457-750 FROM N.A.  
RC TISSUE=Aorta;  
RX MEDLINE=87297534; PubMed=3502711;

RA Tokimatsu I., Tajima S., Nishikawa T., Tajima M., Fukasawa T.;  
RT "Sequence analysis of elastin cDNA from chick aorta and  
RT tissue-specific transcription of the elastin gene in developing chick  
RT embryo";  
RL Arch. Biochem. Biophys. 256:455-461(1987).  
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND  
CC NUCAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.  
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER  
CC INTO AN EXTENSIBLE 3D NETWORK.  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Comment=Additional isoforms seem to exist;  
CC Name=1;  
CC IsoId=P07916-1; Sequence=Displayed;  
CC Name=2; Synonyms=Embryonic;  
CC IsoId=P07916-2; Sequence=VSP 004241, VSP 004242;  
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M18633; AAA48761.1; -;  
DR EMBL; M21880; AAA49082.1; -;  
DR EMBL; M15889; AAA49108.1; -;  
DR PIR; A26601; A26601.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR003979; tropoelastin.  
DR Pfam; PF01391; Collagen; 1.  
DR PRINTS; PR01500; TROPOELASTIN.  
KW Structural protein; Connective tissue; Repeat; Signal;  
KW Alternative splicing.  
FT NON\_TER 1 1  
FT SIGNAL <1 24 ELASTIN.  
FT CHAIN 25 750 8 X TANDEM REPEATS.  
FT DOMAIN 83 686 1.  
FT REPEAT 83 127 2.  
FT REPEAT 219 262 3.  
FT REPEAT 263 318 4.  
FT REPEAT 319 393 5.  
FT REPEAT 394 482 6.  
FT REPEAT 483 554 7.  
FT REPEAT 555 619 8.  
FT REPEAT 620 686 8.  
FT DISULFID 739 745 BY SIMILARITY.  
FT MOD\_RES 63 63 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 66 66 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 111 111 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 115 115 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 156 156 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 159 159 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 198 198 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 200 200 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 235 235 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 252 252 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 256 256 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 297 297 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 301 301 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 354 354 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 357 357 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 427 427 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 431 431 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 513 513 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 517 517 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 520 520 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 586 586 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 590 590 OXIDATIVE DEAMINATION (POTENTIAL).  
FT MOD\_RES 593 593 OXIDATIVE DEAMINATION (POTENTIAL).

```

DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR00139; h aconitase; 1.
DR PROSITE; PS00450; ACONITASE_1; 1.
DR PROSITE; PS01244; ACONITASE_2; 1.
DR Lysine biosynthesis; Lyase; Mitochondrion; Transit peptide;
KW Iron-sulfur.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL) .
FT CHAIN 1 ? 775 HOMOACONITASE.
FT METAL 394 394 IRON-SULFUR (4FE-4S) (BY SIMILARITY) .
FT METAL 462 462 IRON-SULFUR (4FE-4S) (BY SIMILARITY) .
FT METAL 465 465 IRON-SULFUR (4FE-4S) (BY SIMILARITY) .
SQ SEQUENCE 775 AA; 84038 MW; AEC08D6B01169632 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 775;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 347 AVPQTULT 353
Db 63 AVPQTULT 69

RESULT 75
PRZE MYXXA STANDARD; PRT; 777 AA.
AC P18769;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DB DE Gliding motility regulatory protein (EC 2.7.3.-).
GN PRZE.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cystrabacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RP [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=90332690; PubMed=2165608;
RA McCleary W.R., Zusman D.R.;
RT "Prze of Myxococcus xanthus is homologous to both CheA and CheY of
RL Salmonella typhimurium.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:5898-5902(1990).
RN [2]
RN PHOSPHORYLATION OF HIS-49.
RX MEDLINE=91072208; PubMed=2123853;
RA McCleary W.R., Zusman D.R.;
RT "Purification and characterization of the Myxococcus xanthus Prze
RL protein shows that it has autophosphorylation activity.";
RL J. Bacteriol. 172:6661-6668(1990).
CC -1- FUNCTION: PRZE IS INVOLVED IN A SENSORY TRANSDUCTION PATHWAY THAT
CC CONTROLS THE FREQUENCY AT WHICH CELLS REVERSE THEIR GLIDING
CC DIRECTION. PRZE SEEMS TO BE CAPABLE OF AUTOPHOSPHORYLATING ITSELF
CC ON AN HISTIDINE RESIDUE AND THEN TO TRANSFER THAT GROUP TO AN
CC ASPARTATE RESIDUE IN THE C-TERMINAL PART OF THE PROTEIN.
CC -1- SIMILARITY: Contains 1 cheW-like domain.
CC -1- SIMILARITY: Contains 1 histidine kinase domain.
CC -1- SIMILARITY: Contains 1 HPT domain.
CC -1- SIMILARITY: Contains 1 response regulatory domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; M35192; AAA25396.1; --
CC PIR; A35966; A35966.
CC HSP; Q56310; IB3Q.
CC InterPro; IPR003594; ATPbind ATPase.
CC InterPro; IPR004358; Bact_sens_pr_C.
CC InterPro; IPR002545; CheW.

```

```

DR InterPro; IPR005467; His_kinase.
DR InterPro; IPR002570; Hpt.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF01584; Chew; 1.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF01627; Hpt; 1.
DR Pfam; PF00072; response_reg; 1.
DR PRINTS; PR00344; ECTRLSENSOR.
DR ProDom; PD000039; Response_reg; 1.
DR SMART; SM00260; Chew; 1.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00073; HPT; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS00851; CHEW; 1.
DR PROSITE; PS50109; HIS_KIN; 1.
DR PROSITE; PS50894; HPT; 1.
DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW Sensory transduction; Transferrase; Kinase; Phosphorylation.
FT DOMAIN 1 108 HPT.
FT DOMAIN 270 509 HISTIDINE KINASE.
FT DOMAIN 511 645 CHEW-LIKE.
FT DOMAIN 660 776 RESPONSE REGULATORY.
FT MOD_RES 49 49 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT DOMAIN 130 197 ALA/PRO-RICH (POSSIBLE HINGE REGION).
SQ SEQUENCE 777 AA; 83189 MW; 9912BD40991C69E5 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 777;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 PTOAPVA 119
Db 175 PTOAPVA 181

```

Search completed: December 17, 2003, 06:33:48  
 Job time : 15.5265 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:26:06 ; Search time 39.8189 Seconds  
(without alignments)  
2624.667 Million cell updates/sec

Title: US-09-805-550-2

Perfect score: 405

Sequence: 1 MKNVTKLGNFEIERSPD.....BELTANYLLDHGHEFDQQQ 405

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

SPTREMBL\_23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	5.2	378	10 Q94C35	Q94C35 arabidopsis
2	21	5.2	378	10 Q9FF16	Q9FF16 arabidopsis
3	20	4.9	389	10 Q9ST46	Q9ST46 lycopersico
4	18	4.4	113	10 Q9SA20	Q9SA20 arabidopsis
5	18	4.4	365	10 Q94CE9	Q94CE9 arabidopsis
6	18	4.4	371	10 Q8LA46	Q8LA46 arabidopsis
7	17	4.2	382	10 Q03990	Q03990 daucus caro
8	14	3.5	392	10 Q04742	Q04742 oryza sativ
9	13	3.2	246	10 Q9S9L8	Q9S9L8 arabidopsis
10	13	3.2	367	10 Q9MA10	Q9MA10 arabidopsis
11	13	3.2	379	10 Q03991	Q03991 daucus caro
12	13	3.2	419	10 Q9M887	Q9M887 arabidopsis
13	10	2.5	65	10 Q9SCA8	Q9SCA8 lycopersico
14	10	2.5	1860	5 Q9UI29	Q9UI29 leishmania
15	9	2.2	269	5 Q9VI97	Q9VI97 drosophila
16	9	2.2	286	16 Q8P454	Q8P454 xanthomonas

17	9	2.2	365	10 Q9AWS0	Q9AWS0 oryza sativ
18	9	2.2	891	11 Q8VHG2	Q8VHG2 mus musculus
19	9	2.2	3069	16 P95029	P95029 mycobacteri
20	8	2.0	72	2 Q48506	Q48506 lactococcus
21	8	2.0	192	16 Q8D8B5	Q8D8B5 vibrio vuln
22	8	2.0	227	7 Q9TP09	Q9TP09 brachydanio
23	8	2.0	248	16 Q8XRR4	Q8XRR4 ralstonia s
24	8	2.0	254	16 Q8ZB70	Q8ZB70 yersinia pe
25	8	2.0	255	16 Q8DC23	Q8DC23 vibrio vuln
26	8	2.0	256	16 Q9KPF3	Q9KPF3 vibrio chol
27	8	2.0	264	2 Q52007	Q52007 vibrio fisc
28	8	2.0	272	10 Q80422	Q80422 oryza sativ
29	8	2.0	273	10 Q40628	Q40628 oryza sativ
30	8	2.0	278	16 Q8D1B5	Q8D1B5 yersinia pe
31	8	2.0	289	2 Q9X2P4	Q9X2P4 mycobacteri
32	8	2.0	300	10 Q65002	Q65002 betula verr
33	8	2.0	308	10 Q9FUM6	Q9FUM6 betula verr
34	8	2.0	322	7 Q8HWF2	Q8HWF2 brachydanio
35	8	2.0	339	7 Q9GJK0	Q9GJK0 salmo trutt
36	8	2.0	354	5 P91201	P91201 caenorhabdi
37	8	2.0	355	16 Q8PNM4	Q8PNM4 xanthomonas
38	8	2.0	356	16 Q8Y277	Q8Y277 ralstonia s
39	8	2.0	396	9 Q9XJA9	Q9XJA9 streptococc
40	8	2.0	415	10 Q9FNS5	Q9FNS5 chlamydomon
41	8	2.0	425	11 Q9DBZ2	Q9DBZ2 mus musculu
42	8	2.0	462	17 Q9HP89	Q9HP89 halobacteri
43	8	2.0	466	16 Q9RK92	Q9RK92 streptomyce
44	8	2.0	467	10 Q8RV76	Q8RV76 pisum sativ
45	8	2.0	473	16 Q92KE5	Q92KE5 rhizobium m
46	8	2.0	480	10 Q8S0B0	Q8S0B0 oryza sativ
47	8	2.0	514	16 Q8Y1V5	Q8Y1V5 ralstonia s
48	8	2.0	566	16 P72770	P72770 synecocyst
49	8	2.0	571	5 Q9VUD3	Q9VUD3 drosophila
50	8	2.0	595	2 Q85688	Q85688 caulobacter
51	8	2.0	643	16 P72058	P72058 mycobacteri
52	8	2.0	653	16 Q8YGD6	Q8YGD6 brucella me
53	8	2.0	687	11 Q9ABQ3	Q9ABQ3 caulobacter
54	8	2.0	695	11 Q8CI48	Q8CI48 mus musculu
55	8	2.0	704	16 Q8G116	Q8G116 brucella su
56	8	2.0	757	16 Q8VIT6	Q8VIT6 mycobacteri
57	8	2.0	898	11 Q8KLS4	Q8KLS4 mus musculu
58	8	2.0	898	11 Q08721	Q08721 rattus norv
59	8	2.0	908	16 Q8CJX8	Q8CJX8 streptomyce
60	8	2.0	944	11 Q922S3	Q922S3 mus musculu
61	8	2.0	1327	2 Q9X7M2	Q9X7M2 staphylococ
62	8	2.0	1665	11 Q8VII1	Q8VII1 mus musculu
63	8	2.0	1795	16 Q9LCJ9	Q9LCJ9 staphylococ
64	8	2.0	2478	2 Q9RL69	Q9RL69 staphylococ
65	8	2.0	2478	2 Q9LCH2	Q9LCH2 staphylococ
66	8	2.0	2481	16 Q99QR6	Q99QR6 staphylococ
67	8	2.0	2593	12 Q8QL53	Q8QL53 sleeping di
68	8	2.0	2861	5 Q9ULC3	Q9ULC3 leishmania
69	7	1.7	48	6 Q95M44	Q95M44 bos taurus
70	7	1.7	58	16 Q9K8K3	Q9K8K3 bacillus ha
71	7	1.7	74	7 P79662	P79662 oncorhynch
72	7	1.7	74	7 P79652	P79652 oncorhynch
73	7	1.7	74	16 Q8XTY0	Q8XTY0 ralstonia s
74	7	1.7	75	10 Q8L656	Q8L656 oryza sativ
75	7	1.7	83	2 Q9L824	Q9L824 enterococc
76	7	1.7	85	2 Q54454	Q54454 streptococc
77	7	1.7	85	10 Q9XE17	Q9XE17 alexandrium
78	7	1.7	87	16 Q8YV58	Q8YV58 anabaena sp
79	7	1.7	89	15 Q85559	Q85559 baboon endo
80	7	1.7	96	12 Q9QBR1	Q9QBR1 myxoma viru
81	7	1.7	96	12 Q9Q336	Q9Q336 shope fibro
82	7	1.7	100	16 Q99SY0	Q99SY0 staphylococ
83	7	1.7	100	16 Q8CRT7	Q8CRT7 staphylococ
84	7	1.7	105	2 Q8KYV0	Q8KYV0 uncultured
85	7	1.7	107	10 Q8GUR6	Q8GUR6 cyanophora
86	7	1.7	108	16 Q8Z6L4	Q8Z6L4 salmonella
87	7	1.7	108	16 Q84944	Q84944 salmonella
88	7	1.7	109	10 Q8H8Y5	Q8H8Y5 oryza sativ
89	7	1.7	113	2 Q933S0	Q933S0 campylobact



```

Db      269 VOANPQILQPMQLQELGKQNPQ 289
|||||
RESULT 3
Q9STA6 PRELIMINARY; PRT; 389 AA.
AC Q9STA6;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE RAD23 protein.
GN RAD23.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. West Virginia 106; TISSUE=Fruit;
RA Lemaire-Chamley M., Petit J., Raymond P., Chevalier C.;
RT "Analysis of gene expression during early tomato fruit development by
RT mRNA differential display.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ243875; CAB51544.1; -.
DR HSSP; P54725; 1DV0.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STI1.
DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR000626; UBIquitin.
DR Pfam; PF00627; UBA; 2.
DR SMART; SM00727; STI1; 1.
DR SMART; SM00165; UBA; 1.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
SQ SEQUENCE 389 AA; 41508 MW; D2BAEDF0FE70778A CRC64;

Query Match 4.9%; Score 20; DB 10; Length 389;
Best Local Similarity 100.0%; Pred. No. 8.7e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      292 VOANPQILQPMQLQELGKQNP 311
|||||
Db      281 VOANPQILQPMQLQELGKQNP 300
|||||

RESULT 4
Q9SA20 PRELIMINARY; PRT; 113 AA.
AC Q9SA20;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE F309.1 protein.
GN F309.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Vysotskaia V.S., Schwartz J., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J., Li J., Kremenetskaia I., Liu A., Luros J., Gonzalez A.,
RA Altafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
RA Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federpiehl N.A., Theologis A.;
RT "The sequence of BAC F309 from Arabidopsis thaliana chromosome 1.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

```

```

RN      [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006341; AAD34676.1; -.
DR HSSP; P54725; 1DV0.
DR InterPro; IPR000449; UBA_domain.
DR Pfam; PF00627; UBA; 1.
DR SMART; SM00165; UBA; 1.
SQ SEQUENCE 113 AA; 12947 MW; 8BE7CB15C9E12380 CRC64;

Query Match 4.4%; Score 18; DB 10; Length 113;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      295 NPQILQPMQLQELGKQNPQ 312
|||||
Db      5 NPQILQPMQLQELGKQNPQ 22
|||||

RESULT 5
Q94CE9 PRELIMINARY; PRT; 365 AA.
AC Q94CE9;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Putative RAD23 protein.
GN F20B17.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene F20B17.8 (GI:7715605).";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

RN      [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene F20B17.8 (GI:7715605).";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY034912; AAK59419.1; -.
DR EMBL; AY063103; AAL34277.1; -.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STI1.
DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR000626; UBIquitin.
DR Pfam; PF00627; UBA; 2.
DR SMART; SM00727; STI1; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
SQ SEQUENCE 365 AA; 39157 MW; 4PED9EC59B467745 CRC64;

```



Query Match 4.4%; Score 18; DB 10; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 NPQILQPMLOELGKQNPQ 312  
 Db 257 NPQILQPMLOELGKQNPQ 274

## RESULT 6

Q8LA46 PRELIMINARY; PRT; 371 AA.  
 AC Q8LA46;  
 DT 01-OCT-2002 (TREMELrel. 22, Created)  
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE DNA repair protein RAD23, putative.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucotids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,  
 RA Feldmann K.,  
 RA "Full-length cDNA from Arabidopsis thaliana";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY088037; AAM65583.1;  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR006636; STIL.  
 DR InterPro; IPR00449; UBA\_domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS50053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 371 AA; 39747 MW; 081493086EA976E7 CRC64;

Query Match 4.4%; Score 18; DB 10; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-09;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 NPQILQPMLOELGKQNPQ 312  
 Db 263 NPQILQPMLOELGKQNPQ 280

## RESULT 7

O03990 PRELIMINARY; PRT; 382 AA.  
 AC O03990;  
 DT 01-JUL-1997 (TREMELrel. 04, Created)  
 DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE RAD23, isoform I.  
 OS Daucus carota (Carrot).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.  
 OX NCBI\_TaxID=4039;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=W001C;

RX MEDLINE=98345997; PubMed=9681019;  
 RA Sturm A., Leinhard S.;  
 RT "Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in  
 yeast.";  
 RL Plant J. 13:815-821(1998).  
 DR EMBL; Y12013; CAA72741.1; -;  
 DR HSSP; P54725; IDV0.  
 DR InterPro; IPR002965; P-rich\_extensn.  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR006636; STIL.  
 DR InterPro; IPR000449; UBA\_domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR SMART; SMO0727; STIL; 1.  
 DR SMART; SMO0165; UBA; 2.  
 DR SMART; SMO0213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS50053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 382 AA; 40344 MW; 27FF02A2402F3DC2 CRC64;

Query Match 4.2%; Score 17; DB 10; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 NPQILQPMLOELGKQNP 311  
 Db 282 NPQILQPMLOELGKQNP 298

## RESULT 8

Q40742 PRELIMINARY; PRT; 392 AA.  
 AC Q40742;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE OSRAD23.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Nipponbare;  
 RX MEDLINE=97369378; PubMed=9225866;  
 RA Schultz T.F., Quatrano R.S.;  
 RT "Characterization and expression of a rice RAD23 gene.";  
 RL Plant Mol. Biol. 34:557-562(1997).  
 DR EMBL; U63530; AAB65941.1; -;  
 DR HSSP; P54725; IDV0.  
 DR Gramene; Q40742; -;  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR006636; STIL.  
 DR InterPro; IPR000449; UBA\_domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR SMART; SMO0727; STIL; 1.  
 DR SMART; SMO0165; UBA; 2.  
 DR SMART; SMO0213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS50053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 392 AA; 41754 MW; BD6E08574CC7CACB CRC64;

Query Match 3.5%; Score 14; DB 10; Length 392;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 VADVKKRIETTTQGO 36  
 |||||

```

Db      23 VADVKRIIETQQG 36

RESULT 9
Q9S9L8 PRELIMINARY; PRT; 246 AA.
AC Q9S9L8
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE T24D18.27 protein.
GN T24D18.27.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Liu S.-X., Yu G., Sakano H., Jhaveri A., Lee J.M., Lenz C., Pham P.,
RA Toriumi M., Chin C., Chioi J., Choi E., Chung M., Gonzalez A.,
RA Hong B., Koo T., Li J., Liu A., Vaysberg M., Altafi H., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A., Hansen N., Johnson-Hopson C.,
RA Khan S., Kim C., Lam B., Nguyen M., Palm C., Shinn P., Tambunga G.,
RA Khan S., Kim C., Lam B., Nguyen M., Palm C., Shinn P., Tambunga G.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010924; AAF18513.1; -.
DR HSSP; P02248; IUBI.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 1.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00165; UBA; 1.
DR SMART; SM00213; UBQ; 1.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
SQ SEQUENCE 246 AA; 25966 MW; 0867D1E7D72B6FFE CRC64;

Query Match 3.2%; Score 13; DB 10; Length 246;
Best Local Similarity 100.0%; Pred. No. 0.00044;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      182 RALRAAYNNPERA 194
Db      155 RALRAAYNNPERA 167

RESULT 10
Q9WA10 PRELIMINARY; PRT; 367 AA.
ID Q9WA10
AC Q9WA10
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE F20B17.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chioi J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Hong B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,

```

```

RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F20B17 from chromosome
RT I.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chioi J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010793; AAF68123.1; -.
DR HSSP; P54725; 1DV0.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFams; TIGR00601; rad23; 1.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
SQ SEQUENCE 367 AA; 39790 MW; C6D01134B5C69CB6 CRC64;

Query Match 3.2%; Score 13; DB 10; Length 367;
Best Local Similarity 100.0%; Pred. No. 0.00065;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      182 RALRAAYNNPERA 194
Db      168 RALRAAYNNPERA 180

RESULT 11
O03991 PRELIMINARY; PRT; 379 AA.
ID O03991
AC O03991;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE RAD23 protein, isoform II.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W001C;
RA MEDLINE=98345997; PubMed=9681019;
RA Sturm A., Leinhard S.;
RT "Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in
RT yeast.";
RL Plant J. 13:815-821(1998).
DR EMBL; Y12014; CAA72742.1; -.
DR HSSP; P54725; 1DV0.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STI1.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.

```

DR Pfam; PF00240; ubiqtutin; 1.  
 DR SMART; SM00727; STI1; 1.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS0053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 379 AA; 40530 MW; B26697B439CC5929 CRC64;

Query Match 3.2%; Score 13; DB 10; Length 379;  
 Best Local Similarity 100.0%; Pred. No. 0.00067;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 RALRAAYNNPERA 194  
 |||||  
 Db 172 RALRAAYNNPERA 184

## RESULT 12

Q9M887 PRELIMINARY; PRT; 419 AA.  
 AC Q9M887;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative RAD23 (AT3g02540/F16B3\_17).  
 GN F16B3.17.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae;  
 OC eusoids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cnv. Columbia;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,  
 RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
 RA "Arabidopsis thaliana chromosome III BAC F16B3 genomic sequence.";  
 RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Gibson H.A., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,  
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,  
 RA Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis cDNA clones";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Shinn P., Banh J., Bowser L.,  
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,  
 RA Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis ORF clones";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AC021640; AAF32461.1; -;  
 DR EMBL; AV039562; AKG2617.1; -;  
 DR EMBL; AY113034; AAM47342.1; -;  
 DR HSSP; P54725; 1DV0.  
 DR InterPro; IPR002965; P rich\_extensn.  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR000449; UBA\_domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR Pfam; PF00240; ubiqtutin; 1.  
 DR PRINTS; PR01217; PRICEXTENS.

DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS0053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 419 AA; 44247 MW; CAA13BC4FFEB1E25 CRC64;  
 Query Match 3.2%; Score 13; DB 10; Length 419;  
 Best Local Similarity 100.0%; Pred. No. 0.00073;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 TIQQLDMGGGTW 175  
 |||||  
 Db 190 TIQQLDMGGGTW 202

## RESULT 13

Q9SCA8 PRELIMINARY; PRT; 65 AA.  
 AC Q9SCA8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Rad23 protein (Fragment).  
 GN RAD23.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Solanum.  
 OC NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cnv. West Virginia 106; TISSUE=fruit;  
 RA Lemaire-Chamley M., Petit J., Causse M., Raymond P., Chevalier C.;  
 RT "Isolation and characterization of cDNAs expressed during early  
 development of tomato fruit by mRNA differential display";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ270958; CAB65692.1; -;  
 DR HSSP; P54725; 1DV0.  
 DR InterPro; IPR000449; UBA\_domain.  
 DR Pfam; PF00627; UBA; 1.  
 DR SMART; SM00165; UBA; 1.  
 DR NON\_TER  
 FT NON\_TER 1  
 SQ SEQUENCE 65 AA; 7233 MW; 0538945EE903084D CRC64;

Query Match 2.5%; Score 10; DB 10; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 TVTPEERBAI 362  
 |||||  
 Db 16 TVTPEERBAI 25

## RESULT 14

Q9U129 PRELIMINARY; PRT; 1860 AA.  
 AC Q9U129;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Probable kinase.  
 GN L4738.03.  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OC NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Friedlin;  
 RA Hilbert H., Wedler H., Wedler E., Duesterhoef A., Ivens A.C.,  
 RA Murphy L., Quail M., Rajandream M.A., Barrell B.G.;  
 RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Friedlin;  
 RC STRAIN=Friedlin;  
 RX MEDLINE=98146435; PubMed=9477341;

RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
RA Smith D.F.;  
RA "A physical map of the Leishmania major Friedlin genome.";  
RL Genome Res. 8:135-145(1998).  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; AL135930; CAB64574.1; -!  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR Pfam; PF00069; kinase; 1.  
DR ProDom; P0000001; Prot\_kinase; 2.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 1860 AA; 194940 MW; 8843E843429DD7AD CRC64;

Query Match 2.5%; Score 10; DB 5; Length 1860;  
Best Local Similarity 100.0%; Pred.No.2.7; 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0;

Qy 99 PAAPVAPAA 108  
Db 1596 PAAPVAPAA 1605  
|||||

RESULT 15  
Q9V197  
ID Q9V197 PRELIMINARY; PRT; 269 AA.  
AC Q9V197  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CG31496 protein.  
GN NXF4 OR CG14604 OR CG31501.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K.J., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Stapleton M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Dou L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwan C., Jalali M., Kruse D., Li P., Mattel B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradscky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RP Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003672; AAF54028.2; -!  
DR FlyBase; FBGN0051501; nx4.  
SQ SEQUENCE 269 AA; 27072 MW; D69F5D742788F331 CRC64;

Query Match 2.2%; Score 9; DB 5; Length 269;  
Best Local Similarity 100.0%; Pred.No.4.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 PAAPVAPAA 107  
Db 158 PAAPVAPAA 166  
|||||

RESULT 16  
Q8P454  
ID Q8P454 PRELIMINARY; PRT; 286 AA.  
AC Q8P454  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein XCC3862.  
GN XCC3862.  
OS Xanthomonas campestris (pv. campestris).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=340;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33913 / NCPPB 528;  
RX MEDLINE=22022145; PubMed=12024217;

RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Canavan F., Cardoso J., Chambergo F., Ciapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Teai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities";  
 RL Nature 417:459-463 (2002).  
 DR EMBL; AF012507; AAM43093.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 286 AA; 28621 MW; C884041CE7BE6F54 CRC64;

Query Match 2.2%; Score 9; DB 16; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 4.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 PVAPAASVA 110  
 Db 135 PVAPAASVA 143

## RESULT 17

Q9AWS0 PRELIMINARY; PRT; 365 AA.  
 AC Q9AWS0;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Putative DNA binding protein RAV2.  
 GN P0480E02.24.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]\_TaxID=4530;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0480E02";  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF002913; BAB21218.1; -;  
 DR HSSP; O80337; 2GCC.  
 DR Gramene; Q9AWS0; -;  
 DR InterPro; IPR003340; TF\_B3.  
 DR InterPro; IPR001471; TF\_ERF.  
 DR Pfam; PF00847; AP2-domain; 1.  
 DR Pfam; PF02362; B3; 1.  
 DR PRINTS; PR00367; ETRSPLEMMNT.  
 DR ProDom; PD001423; TF\_ERF; 1.  
 DR SMART; SM00380; AP2; 1.  
 SQ SEQUENCE 365 AA; 39280 MW; 0A1EAD4E7986E9E6C CRC64;

Query Match 2.2%; Score 9; DB 10; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 5.6;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTATT 89  
 Db 10 SSGASTATT 18

## RESULT 18

Q8VHG2 PRELIMINARY; PRT; 891 AA.  
 AC Q8VHG2;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Angiotensin.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B6CBAF2; TISSUE=Placenta;  
 RA Troyanovsky B., Bratt A., Holmgren L.;  
 RT "Mouse angiotensin";  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF461135; AAL73436.1; -;  
 DR InterPro; IPR000104; Antifreeze\_1.  
 DR PRINTS; PR00308; ANTIFREEZE1.  
 SQ SEQUENCE 891 AA; 95111 MW; C0544542A4CD6088 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 891;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 PAAPVAPAA 107  
 Db 713 PAAPVAPAA 721

## RESULT 19

P95029 PRELIMINARY; PRT; 3069 AA.  
 AC P95029;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE RV2524C (Fatty-acid synthase).  
 GN FAS OR RV2524C OR MT2600 OR MTCY159.32 OR MTV009.09C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]\_TaxID=1773;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Hoiroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultston J.B., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence";  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z83863; CAB0201.1; -;  
 DR EMBL; AE007096; AAK46907.1; -;

```

DR TIGR; MT2600; -.
DR Tuberculin; RV2524C; -.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR003965; Fatty acid synth.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR002539; MacC dehydratas.
DR Pfam; PF00698; Acyl_transf; 1.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF02801; ketoacyl-synt_C; 1.
DR Pfam; PF01575; MacC dehydratas; 1.
DR PRINTS; PR01483; FASYNTHASE.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
KW Complete proteome.
FT CONFLICT 2699 2699 A -> S (IN REF. 2).
SQ SEQUENCE 3069 AA; 326251 MW; 25EF022D54943D4C CRC64;

Query Match 2.2%; Score 9; DB 16; Length 3069;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 AAPVAPAAAS 108
Db 1754 AAPVAPAAAS 1762
|||||

RESULT 20
Q48506 PRELIMINARY; PRT; 72 AA.
AC Q48506;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Beta-lactamase (Fragment).
GN BETA-LACTAMASE.
OS Lactococcus lactis.
OC Bacteria; Firmicutes; Lactobacilliales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MG1614;
RX MEDLINE=91197101; PubMed=1901704;
RA Sibakov M., Koivula T., von Wright A., Palva I.;
RT "Secretion of TEM beta-lactamase with signal sequences isolated from
the chromosome of Lactococcus lactis subsp. lactis."
RL Appl. Environ. Microbiol. 57:341-348(1991).
DR EMBL; M63293; AAA64243.1; -.
FT NON_TER 72
FT SEQUENCE 72 AA; 7428 MW; 16DFA92111A60D46 CRC64;

Query Match 2.0%; Score 8; DB 2; Length 72;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 75 SKAKASS 82
Db 26 SKAKASS 33
|||||

RESULT 21
Q8DB85 PRELIMINARY; PRT; 192 AA.
AC Q8DB85;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Anthranilate/para-aminobenzoate synthase component II.
GN V13065.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016807; AA011389.1; -.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 192 AA; 20563 MW; B537BF2BABA5946 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 192;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 348 VPQTLTVT 355
Db 140 VPQTLTVT 147
|||||

RESULT 22
Q9TP09 PRELIMINARY; PRT; 227 AA.
AC Q9TP09;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE MHC class I antigen (Fragment).
GN MHC1UDA OR DABE-UDA.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20261668; PubMed=10799891;
RX Michaelova V., Murray B.W., Sultmann H., Klein J.;
RT "A contig map of the Mhc class I genomic region in the zebrafish
reveals ancient synteny."
RL J. Immunol. 164:5296-5305(2000).
CC -I- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
IMMUNE SYSTEM (BY SIMILARITY).
CC -I- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
MICROGLOBULIN) (BY SIMILARITY).
DR EMBL; AF182155; AAF20178.1; -.
DR ZFIN; ZDB-GENE-010116-8; mhcluda.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig-cl.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001039; MHC_I.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00129; MHC_I; 1.
DR ProDom; PD000050; MHC_I; 1.
DR SMART; SM00407; IGC1; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Glycoprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 227
FT SEQUENCE 227 AA; 26616 MW; 22C1B6E02FF81D79 CRC64;

Query Match 2.0%; Score 8; DB 7; Length 227;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 351 TLTVTPEE 358
Db 215 TLTVTPEE 222
|||||

RESULT 23
Q8XRR4 PRELIMINARY; PRT; 248 AA.
AC Q8XRR4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)

```

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Probable transmembrane protein.  
 GN RSP0767 OR RS01941.  
 OS *Ralstonia solanacearum* (*Pseudomonas solanacearum*).  
 OG Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Ralstoniaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GMI1000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brottier P., Camus J.C., Catcolico L.,  
 RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*."  
 RL Nature 415:497-502(2002).  
 DR EMBL: AL646080; CAD17918.1; -;  
 KW Plasmid; Complete proteome.  
 SQ SEQUENCE 248 AA; 24622 MW; 700F48E0F6B9A31B CRC64;

Query Match 2.0%; Score 8; DB 16; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQPAPVA 104  
 DB 94 AQPAPVA 101  
 |||||

RESULT 24  
 Q82BJ0 PRELIMINARY; PRT; 254 AA.  
 AC Q82BJ0;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Pyruvate dehydrogenase complex repressor.  
 GN PDHR OR YPO3420.  
 OS *Yersinia pestis*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebaiha M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Stelton J., Stevens K., Whitehead S., Barrell B.G.;  
 RT "Genome sequence of *Yersinia pestis*, the causative agent of plague."  
 RL Nature 413:523-527(2001).  
 DR EMBL: AJ414157; CAC92650.1; -;  
 DR InterPro; IPR000524; HTH\_GNTR.  
 DR Pfam; PF00392; gntr; 1.  
 DR PRINTS; PR00035; HTHGNTR.  
 DR SMART; SM00345; HTHGNTR; 1.  
 DR PROSITE; PS00043; HTH\_GNTR\_FAMILY; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 254 AA; 29408 MW; 24BACF565816CFE0 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 254;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRL 366  
 DB 52 REAIQRL 59  
 |||||

RESULT 25  
 Q8DC23 PRELIMINARY; PRT; 255 AA.  
 AC Q8DC23;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Pyruvate dehydrogenase complex repressor.  
 GN VV11629.  
 OS *Vibrio vulnificus*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.B.;  
 RT "Complete genome sequence of *Vibrio vulnificus* CMCP6."  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE016802; AAO10048.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 255 AA; 28981 MW; 74965A33E5676FF6 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 255;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRL 366  
 DB 52 REAIQRL 59  
 |||||

RESULT 26  
 Q9KPF3 PRELIMINARY; PRT; 256 AA.  
 AC Q9KPF3;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Pyruvate dehydrogenase complex repressor.  
 GN VC2415.  
 OS *Vibrio cholerae*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=El Tor N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bacs S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*."  
 RL Nature 406:477-483(2000).  
 CC -1- SIMILARITY: BELONGS TO THE GNTR FAMILY OF TRANSCRIPTIONAL REGULATORS.  
 DR EMBL: AE004311; AAF95558.1; -;  
 DR TIGR; VC2415; -;  
 DR InterPro; IPR000524; HTH\_Gntr.  
 DR Pfam; PF00392; gntr; 1.  
 DR PRINTS; PR00035; HTHGNTR.  
 DR SMART; SM00345; HTH\_GNTR; 1.

DR PROSITE; PS00043; HTH\_GNTR\_FAMILY; 1.  
 KW DNA-binding; Transcription regulation; Complete proteome.  
 SQ SEQUENCE 256 AA; 29384 MW; C40833B79005A24B CRC64;

Query Match 2.0%; Score 8; DB 16; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRLE 366  
 |||||  
 DB 52 REAIQRLE 59

## RESULT 27

Q52007 PRELIMINARY; PRT; 264 AA.  
 AC Q52007;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Streptomycin-resistance protein.  
 OS Vibrio fischeri.  
 OG plasmid pRL1063a.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=668;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21519270; PubMed=11607193;  
 RA Wolk C.P., Cai Y., Panoff J.-M.;  
 RT "Use of a transposon with luciferase as a reporter to identify  
 environmentally responsive genes in a cyanobacterium.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5355-5359(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Wolk C.P.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U55385; AA800446.1; -;  
 DR InterPro; IPR006748; APH\_6\_hur.  
 DR Pfam; PF04655; APH\_6\_hur; I.  
 KW Plasmid.  
 SQ SEQUENCE 264 AA; 28853 MW; 92AA3FDDF5BCF745 CRC64;

Query Match 2.0%; Score 8; DB 2; Length 264;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 APAASVAR 111  
 |||||  
 DB 132 APAASVAR 139

## RESULT 28

Q80422 PRELIMINARY; PRT; 272 AA.  
 AC Q80422;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
 DE Carbonic anhydrase.  
 GN P0004A09.27.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Takano T., Liu S.;  
 RT "Cloning of a gene for carbonic anhydrase in rice.";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 clone: P0004A09.";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB016283; BAA31953.1; -;  
 DR EMBL; AP003607; BAB63789.1; -;  
 DR Gramene; O80422; -;  
 DR InterPro; IPR001765; Prok\_COanhhd.  
 DR Pfam; PF00484; Pro\_CA; 1.  
 DR PROSITE; PS00704; PROK\_CO2\_ANNHYDRASE\_1; 1.  
 DR PROSITE; PS00705; PROK\_CO2\_ANNHYDRASE\_2; 1.  
 SQ SEQUENCE 272 AA; 29117 MW; F414E718F846C5D2 CRC64;

Query Match 2.0%; Score 8; DB 10; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 AAPVAPAA 107  
 |||||  
 DB 63 AAPVAPAA 70

## RESULT 29

Q40628 PRELIMINARY; PRT; 273 AA.  
 AC Q40628;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
 DE Carbonic anhydrase.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUR=Leaf;  
 RA Suzuki S., Burnell J.N.;  
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. IR64;  
 RA Wang W., Zhu L.;  
 RT "Isolation and characterization of a gene for carbonic anhydrase in  
 rice.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U08404; AAA86943.1; -;  
 DR EMBL; AF182806; AAD56038.1; -;  
 DR Gramene; Q40628; -;  
 DR InterPro; IPR001765; Prok\_COanhhd.  
 DR Pfam; PF00484; Pro\_CA; 1.  
 DR PROSITE; PS00704; PROK\_CO2\_ANNHYDRASE\_1; 1.  
 DR PROSITE; PS00705; PROK\_CO2\_ANNHYDRASE\_2; 1.  
 SQ SEQUENCE 273 AA; 29204 MW; 686EACB46DD9897A CRC64;

Query Match 2.0%; Score 8; DB 10; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 AAPVAPAA 107  
 |||||  
 DB 64 AAPVAPAA 71

## RESULT 30

Q8D1B5 PRELIMINARY; PRT; 278 AA.  
 AC Q8D1B5;  
 DT 01-MAR-2003 (TReMBLrel. 23, Created)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)



DE Transcriptional regulator for pyruvate dehydrogenase complex.

GN PDHR OR Y0766.

OS Yersinia pestis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Yersinia.

OX NCBI\_TaxID=632;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KIMS / Biovar Mediaevalis;

RX MEDLINE=22137863; PubMed=12142430;

RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,

RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,

RA Retherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,

RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,

RA Perry R.D.;

RT "Genome sequence of Yersinia pestis KIM.";

RL J. Bacteriol. 184:4601-4611(2002).

DR EMBL; AE013678; AA084353.1; --

KW Pyruvate.

SQ SEQUENCE 278 AA; 32020 MW; 9FC639740F8D0CD9 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 278;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 REAIORLE 366

Db 76 REAIORLE 83

RESULT 31

O9X2P4

ID Q9X2P4 PRELIMINARY; PRT; 289 AA.

AC Q9X2P4;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Alkanal monooxygenase-like hypothetical protein.

GN AWLP.

OS Mycobacterium smegmatis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1772;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=mc2155;

RA Kim B.-K., Choi K.-P., Daniels L.;

RT "The gene for f420-dependent glucose-6-phosphate dehydrogenase (fgd)

RT and a conserved upstream gene coding for a beta-lactamase-like protein

RT in Mycobacterium species.";

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF041061; AAD31326.1; --

DR InterPro; IPR002103; Bac\_luciferase.

DR Pfam; PF00296; bac\_luciferase; 1.

KW Hypothetical protein; Monooxygenase.

SQ SEQUENCE 289 AA; 31631 MW; 28776602C9BEFALE CRC64;

Query Match

Best Local Similarity 100.0%; Pred. No. 43;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 267 VVFGAGSG 274

Db 114 VVFGAGSG 121

RESULT 32

O65002

ID O65002 PRELIMINARY; PRT; 300 AA.

AC O65002;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Isoflavone reductase homolog Bet v 6.0101 (Fragment).

GN BERV6.

OS Betula verrucosa (White birch) (Betula pendula).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids I; Fagales; Betulaceae; Betula.

OX NCBI\_TaxID=3505;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pollen;

RX MEDLINE=98178680; PubMed=9519865;

RA Vieths S., Frank E., Scheurer S., Meyer H.E., Hrazdina G.,

RA Hausteiner D.;

RT "Characterization of a new IgB-binding 35-kDa protein from birch

RT pollen with cross-reacting homologues in various plant foods.";

RL Scand. J. Immunol. 47:263-272(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Pollen;

RX MEDLINE=20020186; PubMed=10550744;

RA Karamloo F., Schmitz N., Scheurer S., Foetisch K., Hoffmann A.,

RA Hausteiner D., Vieths S.;

RT "Molecular cloning and characterization of a birch pollen minor

RT allergen, Bet v 5, belonging to a family of isoflavone reductase-

RT related proteins.";

RL J. Allergy Clin. Immunol. 104:991-999(1999).

DR EMBL; AF135127; AAC05116.2; --

DR InterPro; IPR003866; Isoflav\_reduct.

DR Pfam; PF02716; Isoflavone\_redu; 1.

FT NON\_TER 300 300

SQ SEQUENCE 300 AA; 33153 MW; D5B4DEC356E937C2 CRC64;

Query Match

Best Local Similarity 100.0%; Pred. No. 45;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TNFIEAS 18

Db 276 TNFIEAS 283

RESULT 33

O9FUM6

ID O9FUM6 PRELIMINARY; PRT; 308 AA.

AC O9FUM6;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)

DE Allergenic isoflavone reductase-like protein Bet v 6.0102.

GN BERV6.0102.

OS Betula verrucosa (White birch) (Betula pendula).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids I; Fagales; Betulaceae; Betula.

OX NCBI\_TaxID=3505;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pollen;

RX MEDLINE=20020186; PubMed=10550744;

RA Karamloo F., Schmitz N., Scheurer S., Foetisch K., Hoffmann A.,

RA Hausteiner D., Vieths S.;

RT "Molecular cloning and characterization of a birch pollen minor

RT allergen, Bet v 5, belonging to a family of isoflavone reductase-

RT related proteins.";

RL J. Allergy Clin. Immunol. 104:991-999(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Pollen;

RA Karamloo F., Wangorsch A., Hausteiner D., Vieths S.;

RT "Characterization Bet v 6.0102, a high-IgE binding variant of the

RT birch pollen minor allergen, Bet v 6.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF282850; AAG22740.1; --

```

DR InterPro; IPR003866; Isoflav reduct.
DR Pfam; PF02716; Isoflavone reduct; 1.
SQ SEQUENCE 308 AA; 34130 MW; C14AB150751542BC CRC64;

Query Match          2.0%; Score 8; DB 10; Length 308;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      11 TNFEIEAS 18
Db      276 TNFEIEAS 283

RESULT 34
Q8HWF2 PRELIMINARY; PRT; 322 AA.
AC Q8HWF2;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE SI:G2179B20.4 (Major histocompatibility complex class I UDA gene)
DE (Fragment).
GN MHC1UDA.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Hammond S.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL672151; CAD58766.1; -.
FT NON_TER 322 322
SQ SEQUENCE 322 AA; 37121 MW; BD1C4B175DDA90DD CRC64;

Query Match          2.0%; Score 8; DB 7; Length 322;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      351 TLTVTPEE 358
Db      257 TLTVTPEE 264

RESULT 35
Q9GJKO PRELIMINARY; PRT; 339 AA.
AC Q9GJKO;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE MHC class I heavy chain (Fragment).
GN SATR-UBA.
OS Salmo trutta (Brown trout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8032;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-Spleen.
RC MEDLINE=21142794; PubMed=11207285;
EX Shum B.P., Guethlein L., Flodin L.R., Adkison M.A., Hedrick R.P.,
RA Nehring R.B., Stet R.J.M., Secombes C., Parham P.;
RT "Modes of Salmonid MHC Class I and II Evolution Differ from the
RT Primate Paradigm."
RL J. Immunol. 166:3297-3308 (2001).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC IMMUNE SYSTEM (BY SIMILARITY).
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN) (BY SIMILARITY).
DR EMBL; AF296376; AAG02522.1; -.

```

```

DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig ci.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001039; MHC_I.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00129; MHC_I; 1.
DR PRINTS; PR01638; MHCCLASSI.
DR PRODOM; PD000050; MHC_I; 1.
DR SMART; SM00407; IGL1; 1.
DR PROSITE; PS50835; IG LIKE; 1.
KW Glycoprotein; Transmembrane.
FT NON_TER 339 339
SQ SEQUENCE 339 AA; 37983 MW; 67730F33F1E4981D CRC64;

Query Match          2.0%; Score 8; DB 7; Length 339;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      352 LTVTPEER 359
Db      250 LTVTPEER 257

RESULT 36
P91201 PRELIMINARY; PRT; 354 AA.
AC P91201;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical 41.9 kDa protein.
GN EGAP9.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Geisel C.;
RT "The sequence of C. elegans cosmid EGAP9.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U80026; AAC25845.2; -.
DR WormPep; EGAP9.3; CE27912.
DR InterPro; IPR002516; Glyco trans 11.
DR Pfam; PF01531; Glyco_transf_11; 1.
KW Hypothetical protein.
SQ SEQUENCE 354 AA; 41879 MW; E0F60E58CAA1BCDA CRC64;

Query Match          2.0%; Score 8; DB 5; Length 354;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      193 RAIDLYS 200
Db      230 RAIDLYS 237

RESULT 37

```

Q8PNM4 Q8PNM4 PRELIMINARY; PRT; 355 AA.

AC Q8PNM4; DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein XAC1030.

GN XAC1030.

OS Xanthomonas axonopodis (pv. citri).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xanthomonas.

OX NCBI\_TaxID=92829;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=306 / ATCC 13902 / XV 101;

RX MEDLINE=22022145; PubMed=12024217;

RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,

RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,

RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,

RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,

RA Cicatelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,

RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,

RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,

RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,

RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,

RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

RA Spinola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,

RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,

RA Setubal J.C., Kitajima J.P.;

RT "Comparison of the genomes of two Xanthomonas pathogens with differing

RT host specificities";

RL Nature 417:459-463(2002).

DR EMBL; AE011732; AM35913.1; -.

DR InterPro; IPR002965; P rich extensn.

DR PRINTS; PR01217; PRICHEXTENS.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 355 AA; 35642 MW; 195C067A70E75107 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 355;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 APVAPAS 108

Db 270 APVAPAS 277

|||||

RESULT 38

Q8Y277 Q8Y277 PRELIMINARY; PRT; 356 AA.

AC Q8Y277; DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical transmembrane protein RSC0459.

GN RSC0459 OR RS04440.

OS Ralstonia solanacearum (Pseudomonas solanacearum).

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

OC Ralstoniaceae; Ralstonia.

OX NCBI\_TaxID=305;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GM11000;

RX MEDLINE=21681879; PubMed=11823852;

RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,

RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,

RA Chandler W., Choise N., Clauvel-Renard C., Cunnac S., Demange N.,

RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,

RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,

RA Weissenbach J., Boucher C.A.;

RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";

RL Nature 415:497-502(2002).

KW EMBL; AL646059; CAD13987.1; -.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 356 AA; 36540 MW; D468300844371A0C CRC64;

Query Match 2.0%; Score 8; DB 16; Length 356;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQAAPVA 104

Db 176 AQAAPVA 183

|||||

RESULT 39

Q9XJA9 Q9XJA9 PRELIMINARY; PRT; 396 AA.

AC Q9XJA9; DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Putative tail component protein.

OS Streptococcus thermophilus bacteriophage DT1.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.

OX NCBI\_TaxID=90410;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DT1;

RX MEDLINE=99160757; PubMed=10049822;

RA Tremblay D.M., Moineau S.;

RT "Complete genomic sequence of the lytic bacteriophage DT1 of

RT Streptococcus thermophilus";

RL Virology 255:63-76(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=DT1;

RX MEDLINE=21382762; PubMed=11489121;

RA Duplessis M., Moineau S.;

RT "Identification of a genetic determinant responsible for host

RT specificity in Streptococcus thermophilus bacteriophages.";

RL Mol. Microbiol. 41:325-336(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=DT1;

RA Tremblay D.M., Moineau S.;

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF085222; AAD21892.1; -.

DR InterPro; IPR000189; SLT\_domain.

DR Pfam; PF01464; SLT; 1.

SQ SEQUENCE 396 AA; 43254 MW; 3306E7D607E7C137 CRC64;

Query Match 2.0%; Score 8; DB 9; Length 396;

Best Local Similarity 100.0%; Pred. No. 58;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 RDTVVRAL 184

Db 165 RDTVVRAL 172

|||||

RESULT 40

Q9FNS5 Q9FNS5 PRELIMINARY; PRT; 415 AA.

AC Q9FNS5; DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE NADP-malate dehydrogenase (EC 1.1.1.82).

GN NADP-MDH.

OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

OC Chlamydomonadaceae; Chlamydomonas.

OX NCBI\_TaxID=3055;

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=2191;
RC  MEDLINE=21844508; PubMed=11855723; Quesada A.;
RA  Gonzalez I., Mehan P., Fernandez E.,
RT  "NADP-malate dehydrogenase from Chlamydomonas: prediction of new
RL  structural determinants for redox regulation by homology modelling.";
RC  Plant Mol. Biol. 48:211-221(2002).
CC  -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC  -1- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
DR  EMBL; AJ272781; CAC19083.2; --
DR  HSPF; P17606; 7MDH.
DR  InterPro; IPR001236; ldh.
DR  InterPro; IPR001252; Mdh.
DR  Pfam; PF00056; ldh; 1.
DR  Pfam; PF02866; ldh_C; 1.
DR  ProDom; PD003052; Mdh; 1.
DR  PROSITE; PS00068; MDH; 1.
DR  KX  Oxidoreductase; NADP.
SQ  SEQUENCE 415 AA; 44852 MW; 7B32E23DA26D53C8 CRC64;

Query Match 2.0%; Score 8; DB 10; Length 415;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  100 AAVAPAA 107
DB  32 AAVAPAA 39

RESULT 41
ID  Q9DBZ2 PRELIMINARY; PRT; 425 AA.
AC  Q9DBZ2;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE  Speckle-type POZ protein.
GN  SPOB.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C57BL/6J; TISSUE=Lung;
RX  MEDLINE=21085660; PubMed=11217851;
RA  Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA  Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA  Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA  Saito T., Okazaki Y., Gojobori T., Bono H., Kaukawa T., Saito R.,
RA  Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA  Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA  Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA  Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA  Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA  Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA  Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA  Guatinchich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA  Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA  Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA  Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA  Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA  Wynshaw-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohtsuki S.,
RA  Hayashizaki Y.;
RT  "Functional annotation of a full-length mouse cDNA collection.";
RL  Nature 409:685-690(2001).
DR  EMBL; AK004669; BAB23458.1; --
DR  MGB; MGI:1343085; Spop.
DR  InterPro; IPR000210; BTB_POZ.
DR  InterPro; IPR002083; MATH.
DR  Pfam; PF00651; BTB; 1.
DR  Pfam; PF00917; MATH; 1.

```

```

DR  SMART; SM00225; BTB; 1.
DR  SMART; SM00061; MATH; 1.
DR  PROSITE; PS50097; BTB; 1.
SQ  SEQUENCE 425 AA; 47514 MW; 7B6716666C3D884B CRC64;

Query Match 2.0%; Score 8; DB 11; Length 425;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  135 AATVAATD 142
DB  390 AATVAATD 397

RESULT 42
Q9HP89 PRELIMINARY; PRT; 462 AA.
ID  Q9HP89
AC  Q9HP89;
DT  01-MAR-2001 (TrEMBLrel. 16, Created)
DT  01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE  Photolyase/cryptochrome.
GN  PHR1 OR VNG17546.
OS  Halobacterium sp. (strain NRC-1).
OC  Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC  Halobacteriaceae; Halobacterium.
OX  NCBI_TaxID=64091;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=20504483; PubMed=11016950;
RA  Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA  Shukla H.D., Laasy S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA  Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA  Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA  Maddocks D.G., Jablonaki P.E., Krebs M.P., Angevine C.M., Dale H.,
RA  Ikenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA  Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA  Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT  "Genome sequence of Halobacterium species NRC-1";
RL  Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR  EMBL; AE005079; AAG19981.1; --
DR  HSPF; P05327; IQNF.
DR  InterPro; IPR006049; DNA_photolyase.
DR  InterPro; IPR002081; DNA_photolyase_1.
DR  InterPro; IPR006051; DNA_photolyase_C.
DR  InterPro; IPR006050; DNA_photolyase_N.
DR  InterPro; IPR005101; FAD_binding_7.
DR  InterPro; IPR002029; Peptidase_S8.
DR  Pfam; PF00875; DNA_photolyase; 1.
DR  Pfam; PF03441; FAD_binding_7; 1.
DR  PRINTS; PR00147; DNAPHOTLYASE.
DR  ProDom; PD004390; DNA_photolyase_C; 1.
DR  PROSITE; PS00394; DNA_PHOTOLYASE_1; 1.
DR  PROSITE; PS00136; SUBTILASE_ASP; 1.
KW  Lyase; Complete proteome.
SQ  SEQUENCE 462 AA; 51430 MW; 2CB9EE1CFA3976AB CRC64;

Query Match 2.0%; Score 8; DB 17; Length 462;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  181 VRALRAY 188
DB  53 VRALRAY 60

RESULT 43
Q9RK92 PRELIMINARY; PRT; 466 AA.
ID  Q9RK92
AC  Q9RK92;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

```

```

DE Hypothetical protein SC00270.
GN SCO0270 OR SCF1.12
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2)";
RL Nature 417:141-147(2002).
DR EMBL; AL939104; CAB5532.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 466 AA; 49047 MW; 567F8227E4261096 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 466;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 PAATVAAT 141
Db 52 PAATVAAT 59
|||||

RESULT 44
Q8RVT6 PRELIMINARY; PRT; 467 AA.
AC Q8RVT6;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE PSAPV2.
GN PSAPV2.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. midoriisui; TISSUE=Etiolated stem;
RA Kawanara T., Miura A., Kiba A., Toyoda K., Ichinose Y., Shiraiishi T.;
RT "Characterization of Pea Cell Wall-bound Apyrase.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB071370; BAB85978.1; -.
DR InterPro; IPR000407; GDAL_CD39_NTPase.
DR Pfam; PF01150; GDAL_CD39; 1.
DR PROSITE; PS01239; GDAL_CD39_NTPASE; 1.
SQ SEQUENCE 467 AA; 50964 MW; 94167066C46E0D3C CRC64;

Query Match 2.0%; Score 8; DB 10; Length 467;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 KASSSSGAS 85
Db 307 KASSSSGAS 314
|||||

RESULT 45
Q92KE5 PRELIMINARY; PRT; 473 AA.
ID Q92KE5

```

```

AC Q92KE5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein RO1411.
GN RO1411 OR SMC00996.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Porcette D., Puehler A., Fumelle B., Ramepberger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RL Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591787; CAC45990.1; -.
DR InterPro; IPR001967; Ala/AlaCBptase1.
DR Pfam; PF00768; Peptidase S11; 1.
DR PRINTS; PR00725; DADACBPTASE1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 473 AA; 49918 MW; 15044040CA2B6725 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 473;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 VATAETAP 125
Db 372 VATAETAP 379
|||||

RESULT 46
Q8SOB0 PRELIMINARY; PRT; 480 AA.
ID Q8SOB0;
AC Q8SOB0;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Putative RNA helicase.
GN P0470A12.6 OR P0456E05.28.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartioideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone:P0470A12.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone:P0456E05.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
DR EMBL; AP003436; BAB90281.1; -.
DR EMBL; AP003416; BAB92627.1; -.
DR Gramene; Q8SOB0; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000629; DEAD_box.
DR InterPro; IPR001650; Helicase_C.

```

DR Pfam; PF00270; DEAD; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR PROSITE; PS00039; DEAD\_ATP\_HELICASE; 1.  
 KW ATP-binding; Helicase; Hydrolase.  
 SQ SEQUENCE 480 AA; 52191 MW; DB67DB68D6A73994 CRC64;

Query Match 2.0%; Score 8; DB 10; Length 480;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 PAAPVAPA 106  
 |||||  
 Db 46 PAAPVAPA 53

RESULT 47  
 Q8Y1V5 PRELIMINARY; PRT; 514 AA.  
 AC Q8Y1V5;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Probable D-alanyl-D-alanine carboxypeptidase signal peptide  
 DE protein.  
 GN RSC0584 OR RS04879.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Ralstoniaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM1000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Arianoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choiane N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646060; CAD14114.1; -;  
 DR InterPro; IPR000867; Peptidase\_S13.  
 DR Pfam; PF02113; Peptidase\_S13; 1.  
 DR PRINTS; PR00922; DADACBPTASE3.  
 DR TIGRFAMs; TIGR00666; PB4; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 514 AA; 54615 MW; 2356A4830B15EED8 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 AAPATVA 139  
 |||||  
 Db 42 AAPATVA 49

RESULT 48  
 P72770 PRELIMINARY; PRT; 566 AA.  
 AC P72770;  
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE High affinity sulfate transporter.  
 GN SLR1776.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 DR EMBL; D90900; BAA16785.1; -;  
 DR InterPro; IPR002645; STAS.  
 DR InterPro; IPR001902; Sulph\_transpt.  
 DR Pfam; PF01740; STAS; 1.  
 DR Pfam; PF00916; Sulfate transp; 1.  
 DR TIGRFAMs; TIGR00815; sulp; 1.  
 DR PROSITE; PS0801; STAS; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 566 AA; 60928 MW; 11BA8DA1B0FC2FE6 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 257 GVPSSGSN 264  
 |||||  
 Db 155 GVPSSGSN 162

## RESULT 49

Q9VUD3 PRELIMINARY; PRT; 571 AA.  
 AC Q9VUD3;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE CG32139 protein (GH07353p).  
 GN SOX21B OR SOX-LIKE OR CG6419 OR CG13483 OR CG32139.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busch D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Howack J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam J.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RN SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banton J., An H., Baldwin D., Banston J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Hock J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome.";  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.  
RN [3]  
RN SEQUENCE FROM N.A.  
RA Mixa S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome.";  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.  
RN [4]  
RN SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.  
RN [5]  
RN SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.  
RN [6]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AS003535; AAF49754.2; -;  
DR EMBL; AY058313; AAL13542.1; -;  
DR HSSP; Q05066; 1HRX.  
DR FlyBase; FBgn0042630; Sox21b.  
DR InterPro; IPR000910; HMG\_12\_box.  
DR Pfam; PF00505; HMG\_box; 1.  
DR SMART; SM00398; HMG; 1.  
SQ SEQUENCE 571 AA; 60909 MW; 123F3C1C3AACAF29 CRC64;  
Query Match 2.0%; Score 8; DB 5; Length 571;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 341 LGQLAAAV 348  
DB 33 LGQLAAAV 40

## RESULT 50

Q85688 PRELIMINARY; PRT; 595 AA.  
ID AC 085688;  
AC 085688; (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 23, Last annotation update)  
DE Flagellar hook length determination protein homolog.  
DE FLHG.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
OC NCBI\_TaxID=155892;  
OX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CB15;  
RX MEDLINE=89255087; PubMed=2470725;  
RA Mullin D.A., Newton A.;  
RT "Ntr-like promoters and upstream regulatory sequence ftr are required  
for transcription of a developmentally regulated Caulobacter  
crescentus flagellar gene.";  
RT J. Bacteriol. 171:3218-3227(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CB15;  
RA Mullin D.A., Mullin A.H., Ohta N., Newton A.;  
RT "Organization and expression of Caulobacter crescentus genes needed  
for assembly and function of the flagellar hook.";  
RL MOL. Gen. Genet. 265:445-454(2001).  
DR EMBL; AF072135; AAC33327.1; -;  
DR InterPro; IPR000104; Antifreeze\_1.  
DR PRINTS; PR00308; ANTIFREEZE1.  
SQ SEQUENCE 595 AA; 58056 MW; 218BE9BBAA3C93E CRC64;  
Query Match 2.0%; Score 8; DB 2; Length 595;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 115 QAPVATAE 122  
DB 214 QAPVATAE 221

## RESULT 51

P72058 PRELIMINARY; PRT; 643 AA.  
ID AC P72058;  
AC P72058; (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein RV3792.  
GN RV3792 OR MTCY13D12.26.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RL Nature 393:537-544(1998).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

```

CC      -!- SIMILARITY: IN THE C-TERMINAL REGION, TO M.SMEGMATIS KATG.
DR      EMBL; Z80343; CAB02471.1; -.
DR      Hyperculist; Kv3792; -.
KW      Hypothetical protein; Transmembrane; Complete proteome.
FT      TRANSMEM 25 45
FT      TRANSMEM 66 86
FT      TRANSMEM 97 117
FT      TRANSMEM 151 171
FT      TRANSMEM 180 200
FT      TRANSMEM 205 225
FT      TRANSMEM 229 249
FT      TRANSMEM 271 291
FT      TRANSMEM 309 329
FT      TRANSMEM 355 375
FT      TRANSMEM 384 404
FT      TRANSMEM 420 440
FT      TRANSMEM 445 465
FT      TRANSMEM 517 537
FT      TRANSMEM 616 636
SQ      SEQUENCE 643 AA; 69515 MW; 7C58972B085D9EDA CRC64;

Query Match      2.0%; Score 8; DB 16; Length 643;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      341 LGQLAAAV 348
DB      |||||||
        29 LGQLAAAV 36

RESULT 52
Q8YGD6      PRELIMINARY; PRT; 653 AA.
ID      Q8YGD6;
AC      Q8YGD6;
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DT      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE      TRANSPOSASE.
GN      BME11223.
OS      Brucella melitensis.
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC      Brucellaceae; Brucella.
OX      NCBI_TaxID=29459;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=16M / ATCC 23456 / Biotype 1;
RX      MEDLINE=20020109; PubMed=1175668;
RA      DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA      Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA      Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA      Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA      Haselkorn R., Kyripides N., Overbeek R.;
RT      "The genome sequence of the facultative intracellular pathogen
RT      Brucella melitensis.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
DR      EMBL; AE009561; AAL52404.1; -.
DR      InterPro; IPR001584; Rve.
DR      Pfam; PF00665; rve; 1.
KW      Complete proteome.
SQ      SEQUENCE 653 AA; 73916 MW; CB1A023F13D95954 CRC64;

Query Match      2.0%; Score 8; DB 16; Length 653;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      273 SGALDALR 280
DB      |||||||
        255 SGALDALR 262

RESULT 53
Q9ABQ3      PRELIMINARY; PRT; 687 AA.
ID      Q9ABQ3

```

```

AC      Q9ABQ3;
DT      01-JUN-2001 (TrEMBLrel. 17, Created)
DT      01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      General secretion pathway protein D.
GN      CC0173.
OS      Caulobacter crescentus.
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC      Caulobacteraceae; Caulobacter.
OX      NCBI_TaxID=155892;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 19089 / CB15;
RX      MEDLINE=21173698; PubMed=11259647;
RA      Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA      Bisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA      Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA      DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA      Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA      Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA      Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT      "Complete genome sequence of Caulobacter crescentus.";
RL      Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).
DR      EMBL; AE005692; AAK22160.1; -.
DR      TIGR; CC0173; -.
DR      InterPro; IPR001775; Bac_GSPD.
DR      InterPro; IPR004846; GSP1/IIIProtein.
DR      InterPro; IPR005644; NslW-like.
DR      Pfam; PF00263; GSP11_III; 1.
DR      Pfam; PF03958; GSP11_III_N; 3.
DR      PRINTS; PR00811; BCTERIALGSPD.
KW      Complete proteome.
SQ      SEQUENCE 687 AA; 71703 MW; 1EEDBB7E4B11B563 CRC64;

Query Match      2.0%; Score 8; DB 16; Length 687;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      83 GASTATTA 90
DB      |||||||
        307 GASTATTA 314

RESULT 54
Q8CI48      PRELIMINARY; PRT; 695 AA.
ID      Q8CI48
AC      Q8CI48;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Similar to expressed sequence A118201 (Fragment).
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=FVB/N.
RA      Strauberg R.;
RL      Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC037480; AAH37480.1; -.
FT      NON TER 1 1
SQ      SEQUENCE 695 AA; 73412 MW; 46791F8FFA57E5DE CRC64;

Query Match      2.0%; Score 8; DB 11; Length 695;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      134 PAATVAAT 141
DB      |||||||
        7 PAATVAAT 14

```



## RESULT 55

Q8G1I6 PRELIMINARY; PRT; 704 AA.  
AC Q8G1I6;  
DT 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Transposase, putative.  
GN BR0729.  
OS Brucella suis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxID=29461;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1330 / Biovar 1;  
RX MEDLINE=22247741; PubMed=12271122;  
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,  
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,  
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,  
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,  
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.,  
RT "The Brucella suis genome reveals fundamental similarities between  
RT animal and plant pathogens and symbionts."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
DR EMBL; A5014378; AAN29658.1; -.  
DR TIGR; BR0729; -.  
KW Complete proteome.  
SQ SEQUENCE 704 AA; 79768 MW; 6A47F42C372995AD CRC64;

Query Match 2.0%; Score 8; DB 16; Length 704;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 SGALDAR 280

DB 306 SGALDAR 313

## RESULT 56

ID Q8VIT6 PRELIMINARY; PRT; 757 AA.  
AC Q8VIT6;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
DE Hypothetical protein MT3899.1.  
GN MT3899.1.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; A5007183; AAK48265.1; -.  
DR TIGR; MT3899; -.  
KW Hypothetical protein.  
SQ SEQUENCE 757 AA; 82352 MW; 30210944EF4BEAAE CRC64;

Query Match 2.0%; Score 8; DB 16; Length 757;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 LGQLAAAV 348

DB 29 LGQLAAAV 36

## RESULT 57

ID Q8K1S4 PRELIMINARY; PRT; 898 AA.  
AC Q8K1S4;  
DT 01-OCT-2002 (TReMBLrel. 22, Created)  
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Netrin receptor Unc5h1.  
GN UNC5H1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Engelkamp D.;  
RT "Cloning of three mouse unc-5 genes and their expression patterns at  
RT mid-gestation."  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ487852; CAD32250.1; -.  
DR MGI; MGI:894582; Unc5h1.  
DR InterPro; IPR000488; Death.  
DR InterPro; IPR003599; Ig.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003006; IG MHC.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR000906; ZU5.  
DR Pfam; PF00531; death; 1.  
DR Pfam; PF00047; Ig; 1.  
DR Pfam; PF00090; tsp.1; 2.  
DR Pfam; PF00791; ZU5; 1.  
DR SMART; SM00005; DEATH; 1.  
DR SMART; SM00409; IG; 1.  
DR SMART; SM00209; TSP1; 2.  
DR SMART; SM00218; ZU5; 1.  
DR PROSITE; PS50835; IG LIKE; 1.  
DR PROSITE; PS50092; TSP1; 2.  
KW Receptor.  
SQ SEQUENCE 898 AA; 98856 MW; 59F04BA2E196C1DB CRC64;

## Query Match

Best Local Similarity 2.0%; Score 8; DB 11; Length 898;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 LGQLAAAV 348

DB 873 LGQLAAAV 880

## RESULT 58

ID O08721 PRELIMINARY; PRT; 898 AA.  
AC O08721;  
DT 01-JUL-1997 (TReMBLrel. 04, Created)  
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Transmembrane receptor UNC5H1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain, and Ventral spinal cord;  
RX MEDLINE=97271897; PubMed=9126742;  
RA Leonardo E.D., Hinc L., Masu M., Keino-Masu K., Ackerman S.L.,  
RA Tessier-Lavigne M.;  
RT "Vertebrate homologues of C. elegans UNC-5 are candidate netrin

```

RT  receptors.;
RL  Nature 386:833-838(1997).
DR  EMBL; U87305; AAB57678.1; -.
DR  InterPro; IPR000488; Death.
DR  InterPro; IPR003599; IG.
DR  InterPro; IPR007110; IG-like.
DR  InterPro; IPR003006; IG_MHC.
DR  InterPro; IPR000884; TSF1.
DR  InterPro; IPR000906; ZU5.
DR  Pfam; PF00531; death; 1.
DR  Pfam; PF00047; ig; 1.
DR  Pfam; PF00090; LEP_1; 2.
DR  Pfam; PF00791; ZU5_1; 1.
DR  SMART; SM00005; DEATH; 1.
DR  SMART; SM00409; IG; 1.
DR  SMART; SM00209; TSF1; 2.
DR  SMART; SM00218; ZU5; 1.
DR  PROSITE; PS0835; IG_LIKE; 1.
DR  PROSITE; PS0092; TSF1; 2.
KW  Receptor.
SQ  SEQUENCE 898 AA; 98840 MW; 7A3CBCB9E7ACA135 CRC64;

Query Match 2.0%; Score 8; DB 11; Length 898;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 341 LGQLAAAV 348
Db 873 LGQLAAAV 880

RESULT 59
Q8CJKB PRELIMINARY; PRT; 908 AA.
AC Q8CJKB
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Putative transcriptional regulator.
GN SC07173 OR SC8A11.01 OR SC9A4.35.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939130; CAD55526.1; -.
KW Complete proteome.
SQ SEQUENCE 908 AA; 98458 MW; C3D5B12D49623331 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 908;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 342 GQLAAAVP 349
Db 375 GQLAAAVP 382

```

```

RESULT 60
Q922S3 PRELIMINARY; PRT; 944 AA.
AC Q922S3
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DB Unknown (Protein for IMAGE:3594992) (Fragment).
GN A118201.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 4 PDZ/DHR DOMAINS.
DR EMBL; BC006859; AA06859.1; -.
DR MGI; MGI:2145950; A118201.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 4.
DR SMART; SM00228; PDZ; 4.
DR PROSITE; PS0106; PDZ; 4.
FT NON TER 1 1
SQ SEQUENCE 944 AA; 99327 MW; 89CE11BA859B55E4 CRC64;

Query Match 2.0%; Score 8; DB 11; Length 944;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 PAATVAAT 141
Db 256 PAATVAAT 263

RESULT 61
Q9X7M2 PRELIMINARY; PRT; 1327 AA.
AC Q9X7M2
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE CTORF1365, partial (Fragment).
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=99265121; PubMed=10332717;
RA Wu S., de Lencastre H.;
RT "Mrp-a new auxiliary gene essential for optimal expression of
RT methicillin resistance in Staphylococcus aureus.";
RL Microb. Drug Resist. 5:9-18(1999).
DR EMBL; Y09928; CAA71062.1; -.
FT NON TER 1 1
FT NON_TER 1327 1327
SQ SEQUENCE 1327 AA; 140517 MW; B90F2085E800586D CRC64;

Query Match 2.0%; Score 8; DB 2; Length 1327;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 AQAAPVA 104
Db 122 AQAAPVA 129

RESULT 62
Q8VII1 PRELIMINARY; PRT; 1665 AA.
AC Q8VII1

```

```

DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE PDZ-domain protein scribble.
GN A118201.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Cerebellum;
RA Mattock K.L.; Kirschner C.;
RT "Molecular cloning of mouse Scribble cDNA.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 4 PDZ/DHR DOMAINS.
DR EMBL; AF441233; AAL32469.1; -.
DR MGD; MGI:2145950; A118201.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003591; LRR_TYP.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00560; LRR; 13.
DR SMART; SM00369; LRR_TYP; 2.
DR SMART; SM00228; PDZ; 4.
DR PROSITE; PS05056; LRR_TYPICAL; 2.
DR PROSITE; PS0106; PDZ_4.
SQ SEQUENCE 1665 AA; 179749 MW; BE1272F4ECBF010E CRC64;

Query Match      2.0%; Score 8; DB 11; Length 1665;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 PAATVAAT 141
Db 952 PAATVAAT 959

RESULT 63
Q9LCJ9 PRELIMINARY; PRT; 1795 AA.
AC Q9LCJ9;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Truncated FmtB
GN FMTB OR TRUNCATED FMTB OR MW2087.
OS Staphylococcus aureus, and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280, 196620;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KSA8;
RX MEDLINE=20348625; PubMed=10896508;
RA Komatsuzaawa H., Ohta K., Sugai M., Fujiwara T., Glanzmann P.,
RA Berger-Bachi B., Suganaka H.;
RT "Fn551-mediated insertional inactivation of the fmbB gene encoding a
RT cell wall-associated protein abolishes methicillin resistance in
RT Staphylococcus aureus.";
RL J. Antimicrob. Chemother. 45:421-431(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MW2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguni A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
DR EMBL; AB015223; BAA93430.1; -.
DR EMBL; AP004829; BAB95952.1; -.

```

---

```

DR InterPro; IPR00890; Acetate kin.
DR InterPro; IPR005877; Gpos_Ysirk.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
KW Complete proteome.
SQ SEQUENCE 1795 AA; 191008 MW; 6CB77CCFCB33D350 CRC64;

Query Match      2.0%; Score 8; DB 16; Length 1795;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQPAAPVA 104
Db 164 AQPAAPVA 171

RESULT 64
Q9RL69 PRELIMINARY; PRT; 2478 AA.
AC Q9RL69;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Mtp protein.
GN MRP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=97302526; PubMed=9158773;
RA Wu S., de Lencastre H., Sali A., Tomasz A.;
RT "A phosphoglucosaminase-like gene essential for the optimal expression
RT of methicillin resistance in Staphylococcus aureus: molecular cloning
RT and DNA sequencing.";
RL Microb. Drug Resist. 2:277-286(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=97431478; PubMed=9286983;
RA Jolly L., Wu S.W., Van Heijenoort J., de Lencastre H.,
RA Mengin-Lecreulx D., Tomasz A.;
RT "The femR315 gene from Staphylococcus aureus, the interruption of
RT which results in reduced methicillin resistance, encodes a
RT phosphoglucosaminase mutase.";
RL J. Bacteriol. 179:5321-5325(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=99265121; PubMed=10332717;
RA Wu S., de Lencastre H.;
RT "Mip-a new auxiliary gene essential for optimal expression of
RT methicillin resistance in Staphylococcus aureus.";
RL Microb. Drug Resist. 5:9-18(1999).
DR EMBL; Y09927; CAB55329.1; -.
DR InterPro; IPR005877; Gpos_Ysirk.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
SQ SEQUENCE 2478 AA; 263031 MW; 6B9859A02D023C74 CRC64;

Query Match      2.0%; Score 8; DB 2; Length 2478;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQPAAPVA 104
Db 161 AQPAAPVA 168

```

```

RESULT 65
Q9LCH2 PRELIMINARY; PRT; 2478 AA.
AC Q9LCH2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE FmtB.
GN FMTB.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20348625; PubMed=10896508;
RA Komatsuwa H., Ohta K., Sugai M., Fujiwara T., Glanzmann P.,
  Berger-Bachi B., Suginata H.;
RT "Tn51-mediated insertional inactivation of the fmbB gene encoding a
  cell wall-associated protein abolishes methicillin resistance in
  Staphylococcus aureus.";
RL J. Antimicrob. Chemother. 45:421-431(2000).
DR EMBL; AB025716; BAA93438.1; -
DR InterPro; IPR005877; Gpos_YsIRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF04650; YsIRK signal; 1.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR TIGRFAMs; TIGR01168; YsIRK signal; 1.
SQ SEQUENCE 2478 AA; 262995 MW; 1C118EBE0DB03B34 CRC64;

Query Match 2.0%; Score 8; DB 2; Length 2478;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQAAPVA 104
Db 161 AQAAPVA 168

RESULT 66
Q99QR6 PRELIMINARY; PRT; 2481 AA.
AC Q99QR6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE FmtB protein.
GN FMTB (MRP) OR FMTB OR SAV2160 OR SA1964.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
  Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
  Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iino T., Ito T.,
  Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
  Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
  Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
  Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
  Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
  aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003364; BAB58322.1; -
DR EMBL; AP003136; BAB43253.1; -
DR InterPro; IPR000890; Acetate_kin.
DR InterPro; IPR005877; Gpos_YsIRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF04650; YsIRK signal; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.

```

```

DR TIGRFAMs; TIGR01168; YsIRK signal; 1.
DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
SQ COMPLETE PROTEOME.
KW Complete proteome.
SQ SEQUENCE 2481 AA; 263765 MW; E1EAAB99B81665E8 CRC64;

Query Match 2.0%; Score 8; DB 16; Length 2481;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQAAPVA 104
Db 144 AQAAPVA 151

RESULT 67
Q8QL53 PRELIMINARY; PRT; 2593 AA.
AC Q8QL53;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Non-structural polyprotein.
OS Sleeping disease virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
  OC Alphavirus.
OX NCBI_TaxID=78540;
RN [1]
RP SEQUENCE FROM N.A.
RA Bremon M.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Weston J.;
RT "Comparison of two aquatic alphaviruses, Salmon Pancreas Disease Virus
  and Sleeping Disease Virus, using genome sequence analysis, monoclonal
  reactivity and cross-infection.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ316246; CAC87660.1; -
DR InterPro; IPR002589; Alpp.
DR InterPro; IPR002877; FtsJ.
DR InterPro; IPR002620; Peptidase_C9.
DR InterPro; IPR001788; RNA_dep_RNAPol2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR000606; Viral_helicase1.
DR InterPro; IPR002588; V_methyltransf.
DR Pfam; PF01661; Alpp; 1.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF01707; Peptidase_C9; 1.
DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01660; Vmethyltransf; 1.
DR SMART; SM00506; Alpp; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05052; RDRP_VIRAL; 1.
KW Polyprotein.
FT CHAIN 1 563 NON STRUCTURAL PROTEIN P1.
FT CHAIN 564 1420 NON STRUCTURAL PROTEIN P2.
FT CHAIN 1421 1984 NON STRUCTURAL PROTEIN P3.
FT CHAIN 1985 2593 NON STRUCTURAL PROTEIN P4.
SQ SEQUENCE 2593 AA; 284851 MW; 2E2B4F651A45B8CP CRC64;

Query Match 2.0%; Score 8; DB 12; Length 2593;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 APVAPAA 108
Db 1834 APVAPAA 1841

RESULT 68
Q9UIC3

```

```

ID Q9U1C3 PRELIMINARY; PRT; 2861 AA.
AC Q9U1C3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Uncertain, possible lysosomal trafficking regulator or transport
DE protein.
GN L2581.05.
OS Leishmania major.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Zimmermann W., Wambutt R., Ivens A.C., Lawson D., Quail M.,
RA Rajandream M.A., Barrett B.G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL121851; CAB58291.1; -.
DR InterPro; IPR000409; Beige_BEACH.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF02138; Beach; 1.
DR ProDom; PD007848; Beige_BEACH; 1.
DR PROSITE; PS0197; BEACH; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
SQ SEQUENCE 2861 AA; 312361 MW; 2639D42C48AC843C CRC64;

Query Match 2.0%; Score 8; DB 5; Length 2861;
Best Local Similarity 100.0%; Pred. No. 3.9e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 ASSSGAST 86
DB 1547 ASSSGAST 1554

RESULT 69
Q95M44 PRELIMINARY; PRT; 48 AA.
ID Q95M44
AC Q95M44;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Insulin-like growth factor receptor type 2 (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Pfaffl M.W., Mircheva-Georgieva T., Penchev-Georgiev I., Hageleit M.,
RA Blum J.W.;
RT "Real-time RT-PCR quantification of insulin-like growth factor (IGF)-
RT 1, IGF-1 receptor, IGF-2, IGF-2 receptor, insulin receptor, growth
RT hormone receptor, IGF-binding proteins 1, 2 and 3 in the bovine
RT species.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ320234; CAC44343.1; -.
DR InterPro; IPR000479; CIMR.
DR Pfam; PF00878; CIMR; 1.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 48 48
SQ SEQUENCE 48 AA; 5436 MW; 4A27CC2A96E23533 CRC64;

```

```

Query Match 1.7%; Score 7; DB 6; Length 48;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 EGGPGGN 339
DB 38 EGGPGGN 44

RESULT 70
Q9K8K3 PRELIMINARY; PRT; 58 AA.
ID Q9K8K3
AC Q9K8K3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein BH3003.
GN BH3003.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001517; BAB06722.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 58 AA; 6931 MW; 0F04E8CDD8CD6C47 CRC64;

Query Match 1.7%; Score 7; DB 16; Length 58;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 EEREAIQ 363
DB 10 EEREAIQ 16

RESULT 71
P79662 PRELIMINARY; PRT; 74 AA.
ID P79662
AC P79662;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE MHC class I alpha-3 (Fragment).
OS Oncorhynchus tshawytscha (Chinook salmon) (King salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=74940;
RN [1]
RP SEQUENCE OF 3-28 FROM N.A.
RC STRAIN=A3-Onts-HN-11; TISSUE=Liver;
RX MEDLINE=98009322; PubMed=9348703;
RA Miller K.M., Withler R.B., Beacham T.D.;
RT "Molecular evolution at Mhc genes in two populations of chinook salmon
RT Oncorhynchus tshawytscha.";
RL Mol. Ecol. 6:937-954(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3-Onts-HN-11; TISSUE=Liver;
RA Miller K.M.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U80312; AAC78436.1; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_cl.

```

DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00407; IgC1; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 FT NON\_TER 1 74  
 FT NON\_TER 74 74  
 SQ SEQUENCE 74 AA; 8344 MW; A5A5B3481ESB30FE CRC64;

Query Match 1.7%; Score 7; DB 7; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 LVTPEE 358  
 DB 62 LVTPEE 68

## RESULT 72

P79652 ID P79652 PRELIMINARY; PRT; 74 AA.

AC STRAIN-A3-Onts-H-10; TISSUE=Liver;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE MHC class I alpha-3 (fragment).  
 OS Oncorhynchus tshawytscha (Chinook salmon) (King salmon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=74940;  
 RN [1]  
 RP SEQUENCE OF 3-28 FROM N.A.  
 RC STRAIN-A3-Onts-H-10; TISSUE=Liver;  
 RA MEDLINE=98009322; PubMed=9348703;  
 RA Miller K.M., Withler R.E., Beacham T.D.;  
 RT "Molecular evolution at Mhc genes in two populations of chinook salmon  
 Oncorhynchus tshawytscha.";  
 RL Mol. Ecol. 6:937-954(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3-Onts-H-10; TISSUE=Liver;  
 RA Miller K.M.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U80302; AAC78426.1; -.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003597; Ig\_c1.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00407; IgC1; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 FT NON\_TER 1 74  
 FT NON\_TER 74 74  
 SQ SEQUENCE 74 AA; 8320 MW; B1A5B74C1ESB30F9 CRC64;

Query Match 1.7%; Score 7; DB 7; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 LVTPEE 358  
 DB 62 LVTPEE 68

## RESULT 73

QBXTY0 ID QBXTY0 PRELIMINARY; PRT; 74 AA.

AC QBXTY0  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DE Hypothetical protein RSC3415.  
 GN RSC3415 OR RS01795.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Ralstoniaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandelier M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646075; CAD16912.1; -.  
 KW Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 74 AA; 7846 MW; 1C0A9ABF0ECF3758 CRC64;

Query Match 1.7%; Score 7; DB 16; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 ASSSGAS 85  
 DB 2 ASSSGAS 8

## RESULT 74

QB656 ID QB656 PRELIMINARY; PRT; 75 AA.

AC QB656;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN OJ1203D03.2 OR OJ1781E12.2.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,  
 RA Sasaki C., Henry D., Oates R., Simmons J.;  
 RT "Rice Genomic Sequence.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Wang R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Currie J.,  
 RA Collura K.;  
 RT "Rice Genomic Sequence.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC099732; AAM27465.1; -.  
 DR EMBL; AC105927; AAO06954.1; -.  
 DR Gramene; Q8L656; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 75 AA; 8396 MW; F5D1D76BDB3E1228 CRC64;

Query Match 1.7%; Score 7; DB 10; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 AKASSG 83  
 DB 28 AKASSG 34

## RESULT 75

Q9L824 ID Q9L824 PRELIMINARY; PRT; 83 AA.

AC Q9L824;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE Hypothetical 9.6 kDa protein.  
OS Enterococcus faecalis (Streptococcus faecalis).  
OG Plasmid pIP834.  
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
OX NCBI\_TaxID=1351;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BW4382; TRANSPOSON=Tn1549;  
RA Garnier F., Taourit S., Glaser P., Courvalin P., Galimand M.;  
RT "Characterization of transposon Tn1549 conferring VanB-type resistance  
in Enterococcus sp.";  
RL Microbiology 0:0-0(2000).  
DR EMBL; AF192329; AAF72349.1; -.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 83 AA; 9602 MW; 323245C7D9CCEEF7 CRC64;  
  
Query Match 1.7%; Score 7; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 131 QAAPAAT 137  
Db 65 QAAPAAT 71  
|||||  
|||||

Search completed: December 17, 2003, 06:35:14  
Job time : 46.8189 secs





83 7 1.7 65 22 ABG29760 Novel human diageno  
84 7 1.7 66 22 AAU59913 Propionibacterium  
85 7 1.7 66 22 ABP34520 Human ORF3493 prot  
86 7 1.7 71 19 AAW72194 HSV-2 strain SB5 C  
87 7 1.7 80 22 ABG49912 Human liver peptid  
88 7 1.7 80 22 ABG28890 Peptide #2541 enco  
89 7 1.7 80 22 ABB35070 Peptide #2576 enco  
90 7 1.7 80 22 ABB20486 Protein #2485 enco  
91 7 1.7 80 22 AAM55885 Human brain expres  
92 7 1.7 80 22 AAM69258 Human bone marrow  
93 7 1.7 80 22 AAM16081 Peptide #2515 enco  
94 7 1.7 80 22 AAM28574 Peptide #2611 enco  
95 7 1.7 80 22 AAM03808 Peptide #2490 enco  
96 7 1.7 80 23 ABG37793 Human peptid enco  
97 7 1.7 81 22 AAU59644 Propionibacterium  
98 7 1.7 85 16 AAR71469 Premature thermoph  
99 7 1.7 85 21 AAG30527 Arabidopsis thalia  
100 7 1.7 89 22 AAU49184 Propionibacterium

## ALIGNMENTS

## RESULT 1

AA71458  
ID AA71458 standard; Protein; 405 AA.  
AC AA71458;

DT 04-OCT-2000 (first entry)  
DE Maize Rad23 protein #1.

XX Rad23; maize; ATCC No: PTA-530; recombinant expression cassette; wheat;  
KW transgenic plant; soybean; sunflower; sorghum; canola; modulator.  
XX Zea mays.

XX WO200031268-A1.  
XX 02-JUN-2000.

PF 12-OCT-1999; 99WO-US24129.  
XX 23-NOV-1998; 98US-0109728.

XX (PION-) PIONEER HI-BRED INT INC.  
XX Mahajan PB, Tagliani L;

XX WPI; 2000-400078/34.  
XX N-PSDB; AAD01230.

PT Isolated nucleic acid encoding maize RAD23 protein is used to modulate  
PT the levels of polypeptides in plant or in assays for identifying  
PT compounds that bind to and/or increase/decrease enzymatic activity of  
PT catalytically active polypeptides.

XX Claim 11b; Page 75-76; 82pp; English.

XX The present sequence is the maize Rad23 protein #1. It is isolated from  
CC V5 root tissue of a Zea mays cell line B73, infected with corn root  
CC worm. The cDNA sequence is deposited under the ATCC No: PTA-530. Maize  
CC Rad23 DNA sequence operably linked to a promoter can be used to construct  
CC a recombinant expression cassette. This expression cassette can be used  
CC to generate a dicot or monocot transgenic plant e.g., maize, soybean,  
CC sunflower, sorghum, canola, wheat, etc.. It can also be used to modulate  
CC the levels of Rad23 polypeptide expression in a plant or in assays to  
CC identify compounds, that bind to and/or modulate the enzymatic activity  
CC of catalytically active polypeptides.

XX Sequence 405 AA;

Query Match 100.0%; Score 405; DB 21; Length 405;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 M K L N V K T L K G T N F I E A S P D A S V A D V K R I I E T T Q G S T Y R A D Q Q M L I Y Q G K I L K D E T T L E 60  
DB 1 M K L N V K T L K G T N F I E A S P D A S V A D V K R I I E T T Q G S T Y R A D Q Q M L I Y Q G K I L K D E T T L E 60

QY 61 S N G V A E N S F L V I M L S K A S S G A S T A T T A K A P A T L A Q P A P A P A A S V A R T P T Q A P V A T 120  
DB 61 S N G V A E N S F L V I M L S K A S S G A S T A T T A K A P A T L A Q P A P A P A A S V A R T P T Q A P V A T 120

QY 121 A E T A P P S V Q P Q A A P A A T V A A T D D A D V Y S Q A A S N L V F G N N L E O T I Q I L D M G G G T W E R D T V 180  
DB 121 A E T A P P S V Q P Q A A P A A T V A A T D D A D V Y S Q A A S N L V F G N N L E O T I Q I L D M G G G T W E R D T V 180

QY 181 V R A L R A A Y N N P E R A I D Y L Y S G I P E N V E A Q P V A R A P A A G Q Q T N Q A A S P A Q A P A L P V Q P S 240  
DB 181 V R A L R A A Y N N P E R A I D Y L Y S G I P E N V E A Q P V A R A P A A G Q Q T N Q A A S P A Q A P A L P V Q P S 240

QY 241 P A S A G P N A N P L N L P R O G V P S G G S N P G V P F G A G S G A L D A L R Q L P Q F Q A L L Q L V Q A N P Q I L Q 300  
DB 241 P A S A G P N A N P L N L P R O G V P S G G S N P G V P F G A G S G A L D A L R Q L P Q F Q A L L Q L V Q A N P Q I L Q 300

QY 301 P M L Q E L G K N P Q I L R I Q E N Q A E F L R L V N E S P E G G P G G N I L G Q L A A A V P O T L T V T P E E R E 360  
DB 301 P M L Q E L G K N P Q I L R I Q E N Q A E F L R L V N E S P E G G P G G N I L G Q L A A A V P O T L T V T P E E R E 360

QY 361 A T O R L E G M G F N R E L V L E V F F A C N K D E L T A N Y L L D H G H E F D D Q Q Q 405  
DB 361 A T O R L E G M G F N R E L V L E V F F A C N K D E L T A N Y L L D H G H E F D D Q Q Q 405

RESULT 2  
AAG19976  
ID AAG19976 standard; Protein; 307 AA.  
XX AAG19976;  
XX 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21983.  
DE Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 23-APR-1999; 99US-0130891.  
XX 28-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134321.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135153.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140931.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 07-SEP-1999; 99US-0151930.  
PR 10-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.

```

PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match      5.2%; Score 21; DB 21; Length 307;
Best Local Similarity 100.0%; Pred. No. 5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 VQANPOLLOPMLQELGKQNPQ 312
Db 198 VQANPOLLOPMLQELGKQNPQ 218

RESULT 3
AAG45206
ID AAG45206 standard; Protein; 307 AA.
XX
AC AAG45206;
XX
DT 18-OCT-2000 DT (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56725.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.

```



PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158365.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

```

PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      5.2%; Score 21; DB 21; Length 335;
Best Local Similarity 100.0%; Pred. No. 5.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 VQANPOILOPMLQELGKQNPQ 312
Db 226 VQANPOILOPMLQELGKQNPQ 246

RESULT 5
AAG45205
ID AAG45205 standard; Protein; 335 AA.
XX
AC AAG45205;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56724.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000BP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 28-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 06-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.

```

```
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149168.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151108.
PR 31-AUG-1999; 99US-0151130.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155339.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157153.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

QY 292 VQANFQILQPMQLQELGKNPQ 312
Db 226 VQANFQILQPMQLQELGKNPQ 246
RESULT 6
AAG19974
ID AAG19974 standard; Protein; 378 AA.
XX AAG19974;
XX AAG19974;
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21981.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
PF
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
```

Query Match 5.2%; Score 21; DB 21; Length 335;  
Best Local Similarity 100.0%; Pred. No. 5.5e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151085.  
PR 27-AUG-1999; 99US-0151086.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 5.2%; Score 21; DB 21; Length 378;

Best Local Similarity 100.0%; Pred. No. 6.2e-11;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 292 VOANPQILQPMQLQELGKQNPQ 312

Db 269 VOANPQILQPMQLQELGKQNPQ 289

RESULT 7  
AAG45204



ID AAG45204 standard; Protein; 378 AA.  
XX AAG45204;  
AC AAG45204;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56723.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.

```
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 100.0%; Score 21; DB 21; Length 378;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 292 VOANPQILQPMQLQELGKNPQ 312
Db 269 VOANPQILQPMQLQELGKNPQ 289

RESULT 8
AAG44343
ID AAG44343 standard; Protein; 296 AA.
XX
AC
AAG44343;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55533.
```

```
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
```

PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158234.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 4.4%; Score 18; DB 21; Length 296;  
Best Local Similarity 100.0%; Pred. No. 3.2e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 NPQILQPMQLQELGKQNPQ 312  
Db 188 NPQILQPMQLQELGKQNPQ 205

## RESULT 9

AAG17379

ID AAG17379 standard; Protein; 299 AA.

XX

AC AAG17379;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 18374.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0121180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136788.  
PR 01-JUN-1999; 99US-0137228.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137503.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139499.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139762.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145921.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.

```
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.4%; Score 18; DB 21; Length 299;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 NPQLQPMQLQELGKQNPQ 312
DB 191 NPQLQPMQLQELGKQNPQ 208

RESULT 10
AAG44342
ID AAG44342 standard; Protein; 345 AA.
XX AC AAG44342;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 55532.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.

99US-0123180.
99US-0123548.
99US-0125788.
99US-0126264.
99US-0125785.
99US-0127462.
99US-0128234.
99US-0128714.
99US-0129845.
99US-0130077.
99US-0130449.
99US-0130510.
99US-0130891.
99US-0131449.
99US-0132048.
99US-0132407.
99US-0132484.
99US-0132485.
99US-0132486.
99US-0132487.
99US-0132863.
99US-0134256.
99US-0134218.
99US-0134219.
99US-0134221.
99US-0134370.
99US-0134768.
99US-0134941.
99US-0135124.
99US-0135353.
99US-0135629.
99US-0136021.
99US-0136392.
99US-0136782.
99US-0137222.
99US-0137528.
99US-0137502.
99US-0137724.
99US-0138094.
99US-0138540.
99US-0138847.
99US-0139119.
99US-0139452.
99US-0139453.
99US-0139492.
99US-0139454.
99US-0139455.
99US-0139456.
99US-0139457.
99US-0139458.
99US-0139459.
99US-0139460.
99US-0139461.
99US-0139462.
99US-0139463.
99US-0139750.
99US-0139763.
99US-0139817.
99US-0139899.
99US-0140353.
99US-0140354.
99US-0140695.
99US-0140823.
99US-0140991.
99US-0141287.
99US-0141842.
99US-0142154.
99US-0142055.
99US-0142390.
99US-0142803.
99US-0142920.
99US-0142977.
99US-0143542.
```



PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147495.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 4.4%; Score 18; DB 21; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-08;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 NPQILQPMQLQELGKQNPQ 312  
 |||||  
 DB 240 NPQILQPMQLQELGKQNPQ 257

## RESULT 12

AAG44341  
 ID AAG44341 standard; Protein; 368 AA.

XX AAG44341;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55531.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135124.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 17-JUN-1999; 99US-0139453.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144684.  
 PR 21-JUL-1999; 99US-0144814.





PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.

```
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161362.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.4% Score 18; DB 21; Length 371;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 NPQILQPMQLQELGKQNFQ 312
Db 263 NPQILQPMQLQELGKQNFQ 280

RESULT 14
AAG36531
ID AAG36531 standard; Protein; 186 AA.
XX
AC AAG36531;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 44781.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140895.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
```

PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149502.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 21-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 3.2%; Score 13; DB 21; Length 186;  
Best Local Similarity 100.0%; Pred. No. 0.001;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 245 GFNANPLNLFPOG 257  
Db 154 GFNANPLNLFPOG 166  
|||||  
  
RESULT 15  
AAG36530  
ID AAG36530 standard; Protein; 214 AA.  
XX  
AC AAG36530;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 44780.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-01411842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 08-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144894.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145921.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

3.2%; Score 13; DB 21; Length 214;

Best Local Similarity 100.0%; Pred. No. 0.0012;			Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	245	GNANPLNLFPOG 257			
Db	182	GNANPLNLFPOG 194			
RESULT 16					
AAG36529					
ID	AAG36529 standard; Protein; 257 AA.				
XX					
AC	AAG36529;				
DT	18-OCT-2000 (first entry)				
XX					
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 44779.				
XX					
KW	Protein identification; signal transduction pathway; metabolic pathway;				
KW	hybridisation assay; genetic mapping; gene expression control; promoter;				
KW	termination sequence.				
OS	Arabidopsis thaliana.				
XX					
PN	EP1033405-A2.				
XX					
PD	06-SEP-2000.				
XX					
PF	25-FEB-2000; 2000EP-0301439.				
XX					
PR	25-FEB-1999;	99US-0121825.		16-JUN-1999;	99US-0139452.
PR	05-MAR-1999;	99US-0123180.		16-JUN-1999;	99US-0139453.
PR	09-MAR-1999;	99US-0123548.		17-JUN-1999;	99US-0139454.
PR	23-MAR-1999;	99US-0125788.		18-JUN-1999;	99US-0139455.
PR	25-MAR-1999;	99US-0126264.		18-JUN-1999;	99US-0139456.
PR	29-MAR-1999;	99US-0126785.		18-JUN-1999;	99US-0139457.
PR	01-APR-1999;	99US-0127462.		18-JUN-1999;	99US-0139458.
PR	06-APR-1999;	99US-0128234.		18-JUN-1999;	99US-0139459.
PR	08-APR-1999;	99US-0128714.		18-JUN-1999;	99US-0139460.
PR	16-APR-1999;	99US-0129845.		18-JUN-1999;	99US-0139461.
PR	19-APR-1999;	99US-0130077.		18-JUN-1999;	99US-0139462.
PR	21-APR-1999;	99US-0130449.		18-JUN-1999;	99US-0139463.
PR	23-APR-1999;	99US-0130510.		18-JUN-1999;	99US-0139750.
PR	28-APR-1999;	99US-0130891.		18-JUN-1999;	99US-0139751.
PR	28-APR-1999;	99US-0131449.		18-JUN-1999;	99US-0139763.
PR	30-APR-1999;	99US-0132048.		18-JUN-1999;	99US-0139817.
PR	04-MAY-1999;	99US-0132407.		22-JUN-1999;	99US-0139899.
PR	05-MAY-1999;	99US-0132485.		23-JUN-1999;	99US-0140353.
PR	06-MAY-1999;	99US-0132486.		23-JUN-1999;	99US-0140354.
PR	07-MAY-1999;	99US-0132487.		24-JUN-1999;	99US-0140695.
PR	11-MAY-1999;	99US-0132863.		28-JUN-1999;	99US-0140823.
PR	14-MAY-1999;	99US-0134256.		30-JUN-1999;	99US-0140991.
PR	14-MAY-1999;	99US-0134218.		01-JUL-1999;	99US-0141287.
PR	14-MAY-1999;	99US-0134221.		01-JUL-1999;	99US-0141842.
PR	14-MAY-1999;	99US-0134370.		02-JUL-1999;	99US-0142055.
PR	18-MAY-1999;	99US-0134768.		06-JUL-1999;	99US-0142390.
PR	19-MAY-1999;	99US-0134941.		08-JUL-1999;	99US-0142803.
PR	20-MAY-1999;	99US-0135124.		09-JUL-1999;	99US-0142920.
PR	21-MAY-1999;	99US-0135353.		12-JUL-1999;	99US-0142977.
PR	24-MAY-1999;	99US-0135629.		13-JUL-1999;	99US-0143542.
PR	25-MAY-1999;	99US-0136392.		15-JUL-1999;	99US-0143624.
PR	27-MAY-1999;	99US-0136391.		15-JUL-1999;	99US-0144005.
PR	01-JUN-1999;	99US-0137222.		16-JUL-1999;	99US-0144085.
PR	03-JUN-1999;	99US-0137528.		16-JUL-1999;	99US-0144086.
PR	04-JUN-1999;	99US-0137502.		19-JUL-1999;	99US-0144325.
PR	07-JUN-1999;	99US-0137724.		19-JUL-1999;	99US-0144331.
PR	08-JUN-1999;	99US-0138094.		19-JUL-1999;	99US-0144332.
PR	10-JUN-1999;	99US-0138540.		19-JUL-1999;	99US-0144334.
PR	10-JUN-1999;	99US-0138847.		20-JUL-1999;	99US-0144335.
PR	14-JUN-1999;	99US-0139119.		20-JUL-1999;	99US-0144632.

PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149302.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150366.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3.2%; Score 13; DB 21; Length 257;  
Best Local Similarity 100.0%; Pred. No. 0.0014;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 245 GPNANPLNLPQ 257  
Db 225 GPNANPLNLPQ 237

RESULT 17  
AAY71459

ID AAY71459 standard; Protein; 368 AA.

XX AAY71459;

DT 04-OCT-2000 (first entry)

DE Maize Rad23 protein #2.

KW Rad23; maize; ATCC No: PTA-531; recombinant expression cassette; wheat;  
transgenic plant; soybean; sunflower; sorghum; canola; modulator.

OS Zea mays.

PN WO200031268-A1.

PD 02-JUN-2000.

PF 12-OCT-1999; 99WO-US24129.

PR 23-NOV-1998; 98US-0109728.

PA (PION-) PIONEER HI-BRED INT INC.

PI Mahajan PB, Tagliani L;

DR WPI; 2000-400078/34.

DR N-PSDB; AAD01231.

PT Isolated nucleic acid encoding maize RAD23 protein is used to modulate  
the levels of polypeptides in plant or in assays for identifying  
compounds that bind to and/or increase/decrease enzymatic activity of  
catalytically active polypeptides -

PS Claim 11b; Page 78-79; 82pp; English.

CC The present sequence is the maize Rad23 protein #2. It is isolated from  
a Zea mays cell line, B73 callus tissue regenerated five days after  
transfer of the callus from medium containing auxin to a medium devoid  
of exogenous auxin. The cDNA is deposited under the ATCC No: PTA-531.  
CC Maize Rad23 DNA sequence operably linked to a promoter, can be used to  
construct a recombinant expression cassette. This expression cassette  
can be used to generate a dicot or monocot transgenic plant e.g., maize,  
soybean, sunflower, sorghum, canola, wheat, etc.. It can also be used to  
modulate the levels of Rad23 polypeptide expression in a plant or in  
assays to identify compounds, that bind to and/or modulate the enzymatic  
activity of catalytically active polypeptides.

SQ Sequence 368 AA;

Query Match 3.2%; Score 13; DB 21; Length 368;  
Best Local Similarity 100.0%; Pred. No. 0.002;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 182 RALRAAYNNPERA 194

Db 167 RALRAAYNNPERA 179

RESULT 18  
ABP34618

ID ABP34618 standard; Protein; 117 AA.

XX ABP34618;

DT 08-JUL-2002 (first entry)

DE Human ORF3591 protein, SEQ ID NO:7182.

KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
disease monitoring; cytokine; cell proliferation; cell differentiation;

immune modulation; haematopoiesis regulation; tissue growth; angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic; thrombolytic; tumour inhibition; bodily characteristic; fertility; behaviour; cancer; proliferative disorder; neurological disorder; cardiovascular disease; immune system disorder; organ transplantation; tissue growth disorder; tissue regeneration disorder; diabetes mellitus; hypothyroidism; cholesterol ester storage disease; infection; vulnery; vasotropic; antipruritic; antidiabetic; cytostatic; neurotropic; neuroprotective; antithrombotic; anticoagulant; thrombolytic; cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator; dermatological; analgesic; virucide; antibacterial; fungicide.

Homo sapiens.

WO200190366-A2.

29-NOV-2001.

24-MAY-2001; 2001WO-US17076.

24-MAY-2000; 2000US-206690P.

(CURA-) CURAGEN CORP.

Leach MD, Shinkets RA;

WPI; 2002-106200/14.

N-PSDB; ABN78644.

Novel human polypeptides and polynucleotides useful for diagnosing, preventing and treating cardiovascular disease, neurodegenerative, hyperproliferative disorders and disorders related to organ transplantation -

Claim 10; Page 2042; 2508pp; English.

Sequences ABP31028-ABP3561 represent 4534 novel human proteins designated ORF (open reading frame) 1-4534, and sequences ABN75054-ABN79587 represent cDNAs encoding them. The invention also encompasses polypeptides at least 80% identical to the ORF1-ORF4534 (collectively referred to as ORFX) proteins, polynucleotides at least 85% identical to the ORFX nucleic acid sequences, vectors and host cells comprising ORFX polynucleotides, the recombinant production of ORFX proteins, antibodies specific for ORFX proteins, methods of detecting ORFX polynucleotides and activity, and methods of screening for modulators of ORFX expression or ORFX-associated disorder. The ORFX proteins of the invention have a wide range of biological activities, such as cytokines, cell proliferation, cell differentiation, immune modulation, haematopoiesis regulation, tissue growth, angiogenesis, activin or inhibin activity, chemotactic/chemokinetic activity, haemostatic activity, thrombolytic activity, receptor/ligand, antiinflammatory activity, tumour inhibition activity, and antiinfective activity, and may also be involved in the determination of bodily characteristics, fertility and behaviour. ORFX proteins, nucleic acids and antibodies may be used in the treatment of cancers, other proliferative disorders such as psoriasis and benign tumours, neurological disorders such as epilepsy and Alzheimer's disease, cardiovascular diseases, immune system disorders, disorders related to organ transplantation, disorders of tissue growth and regeneration, diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester storage disease, and infectious diseases caused by viral, bacterial, fungal and other pathogens. ORFX nucleic acids may also be used as a source of primers and probes, in the detection of ORFX genomic sequences or transcripts, in the identification and cloning of homologous sequences, in genetic diagnosis, and in forensic biology. The ORFX nucleic acids may additionally be used to produce transgenic animals which may be useful for studying the function and/or activity of ORFX protein, and in drug screening. The ORFX proteins may also be used as immunogens to generate specific antibodies, which are useful in the diagnosis, treatment and monitoring of ORFX-associated diseases.

Sequence 117 AA;

Query Match 2.5%; Score 10; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 0.43;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KQNPOILRLI 317  
DB 51 KQNPOILRLI 60  
|||||

RESULT 19

ABB69580  
ID ABB69580 standard; Protein; 487 AA.

XX ABB69580;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 35532.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL13683.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX Disclosure; SEQ ID NO 35532; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 487 AA;

Query Match 2.2%; Score 9; DB 22; Length 487;

Best Local Similarity 100.0%; Pred. No. 15;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 PAAPVAPAA 107  
DB 158 PAAPVAPAA 166  
|||||

RESULT 20

AAG81195  
ID AAG81195 standard; Protein; 3070 AA.

XX



AC AAG81195;  
 XX  
 DT 04-SEP-2001 (first entry)  
 XX  
 DE Mycobacterium tuberculosis potential drug target protein SEQ ID 246.  
 XX  
 KW Drug target; growth; organism viability; characterisation.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO200135317-A1.  
 XX  
 PD 17-MAY-2001.  
 XX  
 XX 13-NOV-2000; 2000WO-US311152.  
 XX  
 PR 12-NOV-1999; 99US-0165086.  
 XX  
 PR 12-NOV-1999; 99US-0165124.  
 XX  
 PR 01-FEB-2000; 2000US-0179531.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Eisenberg D, Rotstein SH, Marcotte EM;  
 XX  
 XX N-PSDB; AAH52046.  
 XX  
 XX Identifying nucleotide or polypeptide sequence for use as drug target,  
 XX involves providing algorithm that analyzes a functional relationship  
 XX between nucleotide or polypeptide sequences, and comparing the  
 XX sequences -  
 XX  
 XX Disclosure; Page 178-180; 207pp; English.  
 XX  
 XX This invention relates to a method for identifying a nucleotide or  
 XX polypeptide sequence that may be a drug target, or essential for growth  
 XX or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
 XX represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium  
 XX tuberculosis proteins which are potential drug targets. The DNA and  
 XX protein sequences are used to illustrate the method of the invention. The  
 XX method involves providing an unknown nucleotide or polypeptide sequences,  
 XX and comparing it to a number of sequences along with at least one  
 XX algorithm capable of analysing a functional relationship between  
 XX nucleotide and polypeptide sequences. The method is useful for  
 XX characterising the function of nucleic acids and polypeptides that may be  
 XX useful as a target for a drug or essential for the growth or viability of  
 XX an organism.  
 XX  
 XX SQ Sequence 3070 AA;  
 XX  
 XX Query Match 2.2%; Score 9; DB 22; Length 3070;  
 XX Best Local Similarity 100.0%; Pred. No. 94;  
 XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX QY 100 AAPVAPPAAS 108  
 XX |||||  
 XX 1755 AAPVAPPAAS 1763  
 XX  
 XX RESULT 21  
 XX AAR74046  
 XX ID AAR74046 standard; Peptide; 23 AA.  
 XX  
 XX AC AAR74046;  
 XX  
 XX XX 25-MAR-2003 (updated)  
 XX DT 01-DEC-1995 (first entry)  
 XX  
 XX XX Rice carbonic-anhydrase.  
 XX  
 XX XX Rice; carbonic-anhydrase; carbon-fixation; monocotyledon; cereal;  
 XX crop improvement.  
 XX

OS Oryza sativa.  
 XX  
 XX Key Location/Qualifiers  
 XX FT Misc-difference 17 /note= "amino acid at position 17 is not identified  
 XX FT in the specification"  
 XX  
 XX PN WO9511979-A1.  
 XX  
 XX PD 04-MAY-1995.  
 XX  
 XX PF 27-OCT-1994; 94WO-JP01814.  
 XX  
 XX PR 29-OCT-1993; 93JP-0294278.  
 XX  
 XX XX (NISR ) JAPAN TOBACCO INC.  
 XX  
 XX PI Burnell NJ, Suzuki S;  
 XX  
 XX DR WPI; 1995-178871/23.  
 XX  
 XX Cloned DNA coding for monocotyledon carbonic anhydrase - is used  
 XX PT to improve carbon fixation in monocotyledonous plants.  
 XX  
 XX PS Disclosure; Page 17; 42pp; Japanese.  
 XX  
 XX CC The cDNA sequence given in AAQ92067 was obt'd. by screening and  
 XX cloning a cDNA library generated from rice RNA, and encodes  
 XX carbonic-anhydrase (AAR74043). A peptide corresp. to amino acids  
 XX 63-85 of the encoded enzyme is given in AAR74046.  
 XX (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 XX SQ Sequence 23 AA;  
 XX  
 XX Query Match 2.0%; Score 8; DB 16; Length 23;  
 XX Best Local Similarity 100.0%; Pred. No. 6.7;  
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX QY 100 AAPVAPPA 107  
 XX |||||  
 XX 1 AAPVAPPA 8  
 XX  
 XX DB  
 XX  
 XX RESULT 22  
 XX AEG12648  
 XX ID AEG12648 standard; Protein; 71 AA.  
 XX  
 XX AC AEG12648;  
 XX  
 XX XX 18-FEB-2002 (first entry)  
 XX  
 XX DE Novel human diagnostic protein #12639.  
 XX  
 XX XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 XX OS Homo sapiens.  
 XX  
 XX PN WO200175067-A2.  
 XX  
 XX PD 11-OCT-2001.  
 XX  
 XX PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX  
 XX XX (HYSE-) HYSEQ INC.  
 XX  
 XX XX Drmanac RT, Liu C, Tang YT;  
 XX  
 XX WPI; 2001-639362/73.  
 XX  
 XX DR N-PSDB; AAS76835.  
 XX

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 43007; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 71 AA;

Query Match 2.0%; Score 8; DB 22; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 AAPVAPAA 107  
 Db 35 AAPVAPAA 42  
 |||||

RESULT 23  
 AAR99709  
 ID AAR99709 standard; Protein; 78 AA.  
 AC AAR99709;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 26-SEP-1996 (first entry)  
 XX  
 XX pKTH1798-encoded protein.  
 XX  
 KW Promoter; signal sequence; protein secretion; vector;  
 KW Gram-positive bacterium; Escherichia coli; Lactobacillus;  
 KW Bacillus subtilis; pKTH1798; beta-lactamase.  
 XX  
 OS Lactococcus lactis subsp. lactis.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 77..78  
 FT /note= "N-terminus of TEM-beta-lactamase"  
 XX  
 XX US5529908-A.  
 XX  
 XX 25-JUN-1996.  
 PD  
 XX 10-FEB-1993; 93US-0015582.  
 PF  
 XX 10-JUL-1989; 89US-0377450.  
 PR 10-FEB-1993; 93US-0015582.  
 XX  
 XX (VALI-) VALIO LTD.  
 PA  
 XX

PI Koivula T, Palva I, Sibakov M, Von Wright A;  
 XX WPI; 1996-308737/31.  
 DR N-PSDB; AAT31878.  
 XX  
 PT Lactococcus promoters and signal sequences - and related vectors and  
 PT transformed cells, provide efficient expression of proteins in Gram  
 PT positive bacteria  
 XX  
 XX Example 5; Fig 15; 20pp; English.  
 XX  
 CC Lactococcus lactis subsp. lactis-derived plasmid pKTH1798 (AAT31878)  
 CC codes for a protein (AAR99709) that includes the N-terminal region of  
 CC the TEM-beta-lactamase gene. The correct reading frame of pKTH1798  
 CC was detd. by matching the 3 reading frames with the known reading  
 CC frame of beta-lactamase. pKTH1798 promoter and signal sequences may  
 CC be incorporated into plasmids and used to achieve enhanced  
 CC heterologous gene expression and protein secretion in E. coli and  
 CC Gram-positive bacteria, esp. Bacillus subtilis, Lactococcus and  
 CC Lactobacillus spp.  
 CC (Updated on 25-MAR-2003 to correct PF field.)  
 XX  
 SQ Sequence 78 AA;  
 Query Match 2.0%; Score 8; DB 17; Length 78;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 75 SKAKASSS 82  
 Db 33 SKAKASSS 40  
 |||||

RESULT 24  
 AAR22951  
 ID AAR22951 standard; Protein; 117 AA.  
 AC AAR22951;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 15-OCT-1992 (first entry)  
 XX  
 XX Promoter/secretion promoting signal sequence from pKTH1797.  
 XX  
 KW Probe-vector; secretion; signal; promoter; plasmid; expression;  
 KW heterologous; homologous; E. coli; Gram-positive bacteria;  
 KW TEM; beta-lactamase; TEM-bla.  
 XX  
 OS Lactobacillus lactis subsp. lactis.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 116..117  
 FT /note= "N-terminus of TEM-bla"  
 XX  
 XX WO9204451-A.  
 XX  
 XX 19-MAR-1992.  
 PD  
 XX 30-AUG-1990; 90WO-FI00204.  
 PF  
 XX 30-AUG-1990; 90WO-FI00204.  
 PR  
 XX (GENE-) GENESIT OY.  
 PA (VALI-) VALIO FINNISH COOPERATIVE DAIRIES ASSOC.  
 XX  
 XX Palva I, Sibakov M, Koivula T, Vonwright A, Sibakov MY;  
 PI WPI; 1992-131815/16.  
 DR N-PSDB; AAQ23883.  
 XX  
 XX Promoter probe vectors for protein expression - replicable in E.  
 PT coli, B. subtilis, Lactococci and Lactobacillus  
 PT  
 XX

PS Disclosure; Fig 14; 70pp; English.

XX The three reading frames of the sequence represented in AAQ23883 were  
CC matched with the known reading frame of beta-lactamase, in order to  
CC determine the correct reading frame. The deduced amino acid  
CC sequence of one reading frame is given here, and is an exact  
CC reproduction of the sequence given in the specification, i.e.  
CC including the seven X's which are encoded by stop codons.  
CC Promoter probe-vectors pKTH1734, pKTH1736 and pKTH1750 were used to  
CC clone and sequence previously unknown and undescribed L. lactis subsp.  
CC lactis promoter and promoter/secretion signal promoting nucleotide  
CC sequences. These sequences are contained in pKTH1816 (AAQ23878),  
CC pKTH1817 (AAQ23879), pKTH1820 (AAQ23880), pKTH1874 (AAQ23881), pKTH1789  
CC (AAQ23882), pKTH1797 (AAQ23883), pKTH1798 (AAQ23884), pKTH1799  
CC (AAQ23885), pKTH1801 (AAQ23886), pKTH1821 (AAQ23887). The sequences may  
CC be beneficially incorporated into plasmids, by means of which it is  
CC possible to achieve enhanced heterologous (and homologous) protein  
CC expression in E. coli and, esp., in Gram-positive bacteria.  
CC (Updated on 25-MAR-2003 to correct PF field.)  
CC (Updated on 25-MAR-2003 to correct PR field.)  
CC (Updated on 25-MAR-2003 to correct PA field.)  
CC (Updated on 25-MAR-2003 to correct PI field.)  
XX

SQ Sequence 117 AA;

Query Match 2.0%; Score 8; DB 13; Length 117;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 SKAKASSS 82

DB 72 SKAKASSS 79

RESULT 25

AA47659

ID AAR47659 standard; Protein; 117 AA.

AC AAR47659;

DT 25-MAR-2003 (updated)

DT 07-MAR-1994 (first entry)

XX Promoter-secretion signal peptide from *Lactococcus lactis*.

XX Promoter; secretion; heterologous proteins; homologous proteins;  
XX expression cassette; *Lactococcus lactis*.

OS *Lactococcus lactis*.

XX US5242821-A.

XX 07-SEP-1993.

XX 10-JUL-1989; 89US-0377450.

XX 10-JUL-1989; 89US-0377450.

XX (VALI-) VALIO.

XX Palva I, Sibakov M, Koivula T, Von Wright A;

XX WPI; 1993-295228/37.

XX N-PSDB; AAQ46208.

XX New promoters and secretion promoting signals - isolated from  
XX *Lactococcus lactis*, used for expression of proteins in  
XX Gram-positive bacteria

XX Disclosure; Claim 4; Figure 15; 32pp; English.

XX The *Lactococcus lactis* promoter-secretion signal coding sequence is  
XX removed from the plasmid pKTH1798 and is operably linked to a

CC *Lactococcus lactis* promoter sequence. By using L. lactis promoters  
CC and promoter/secretion signals, the production of heterologous and  
CC homologous proteins in E.coli and gram-positive hosts can be greatly  
CC improved.  
CC (Updated on 25-MAR-2003 to correct PI field.)  
XX

SQ Sequence 117 AA;

Query Match 2.0%; Score 8; DB 14; Length 117;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 SKAKASSS 82

DB 72 SKAKASSS 79

RESULT 26

AA26036

ID AAY26036 standard; Protein; 186 AA.

AC AAY26036;

DT 22-OCT-1999 (first entry)

XX Secreted protein if87\_1.

XX Secreted protein; cytokine; cell proliferation; immune stimulation;  
XX vaccine; immune suppression; haematopoiesis; tissue growth; activin;  
XX inhibitor; chemotaxis; chemokinesis; haemostasis; thrombolytic;  
XX receptor; ligand; anti-inflammatory; cadherin; tumour; gene therapy.

OS *Homo sapiens*.

XX WO9928335-A1.

XX 10-JUN-1999.

XX 02-DEC-1998; 98WO-US25512.

XX 30-NOV-1998; 98US-0203106.

XX 04-DEC-1997; 97US-0067454.

XX (GENY) GENETICS INST INC.

XX Agostino MJ, Clark HF, Collins-Racie LA, Evans C;

XX Fecthel K, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX Steininger RJ, Treacy M, Wong GG;

XX WPI; 1999-385352/32.

XX N-PSDB; AAX80671.

XX New polynucleotides encoding secreted human proteins

XX Claim 25; Page 111; 124pp; English.

XX The present sequence is a secreted protein if87\_1, encoded by known clone  
XX if87\_1 (deposited as ATCC 98600) isolated from human adult uterus CDNA  
XX library. Recombinant secreted proteins can be produced by  
XX transforming host cells and culturing them under suitable conditions.  
XX The polynucleotide and protein are predicted to have biological  
XX activities which would make them suitable for treating, preventing or  
XX ameliorating medical conditions in humans and animals. Some predicted  
XX biological activities include cytokine and cell proliferation/  
XX differentiation activity, immune stimulating (e.g. as vaccines) or  
XX suppressing activity, haematopoiesis regulating activity, tissue growth  
XX activity, activin/inhibin activity, chemotactic/chemokinetic activity,  
XX hemostatic and thrombolytic activity, receptor/ligand activity, anti-  
XX inflammatory activity, cadherin/tumour invasion suppressor activity, and  
XX tumour inhibition activity. The polynucleotide encoding secreted  
XX protein can be used for gene therapy.

XX Sequence 186 AA;

Query Match 2.0%; Score 8; DB 20; Length 186;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
 Db 39 SSGASTAT 46

RESULT 27  
 AAU39003  
 ID AAU39003 standard; Protein; 186 AA.  
 XX AC AAU39003;  
 XX DT 16-JAN-2002 (first entry)  
 XX DE Human secreted protein if87\_1.  
 XX KW Human; secreted protein; antiinflammatory; immunosuppressive;  
 KW nontropic; neuroprotective; antiarthritic; antimicrobial; vulneryary;  
 KW cytosolic; antidiabetic; virucide; antifertility; anticonvulsant;  
 KW vasotropic; antiparkinsonian; immunostimulant; dermatological;  
 KW antirheumatic; antitumor; antitumor; osteopathic; tranquiliser;  
 KW cerebroprotective; cytokine; cell proliferation; cell differentiation;  
 KW immune deficiency; severe combined immunodeficiency; SCID; tumour;  
 KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;  
 KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;  
 KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; infection; cardiac disease;  
 KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;  
 KW food supplement; vaccine.  
 XX OS Homo sapiens.  
 XX KW WO200175068-A2.  
 XX FN 11-OCT-2001.  
 XX PD 22-MAR-2001; 2001WO-US09369.  
 XX PF 30-MAR-2000; 2000US-0539330.  
 XX PR 04-DEC-2000; 2000US-0729674.  
 XX (GEM) GENETICS INST INC.  
 XX PA Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;  
 PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;  
 PI Clark H, Fechtel K, Merberg D;  
 XX WPI; 2001-639363/73.  
 XX DR N-PSDB; AAS59221.  
 XX Secreted human proteins, useful as vaccine for treating various  
 PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and  
 PT nervous system disorders (e.g. stroke) -  
 XX Disclosure; Page 476; 619pp; English.  
 XX The invention relates to novel human secreted proteins, the nucleic  
 CC acids encoding them. The protein may exhibit cytokine, cell proliferation  
 CC or cell differentiation activity or may induce production of other  
 CC cytokines in certain cell populations and may exhibit immune stimulating  
 CC or immune suppressing activity, which is useful for the treatment of  
 CC various immune deficiencies and disorders e.g. severe combined  
 CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,  
 CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
 CC inflammation. The proteins are also useful in the treatment of diseases  
 CC and disorders including tissue, skin and organ transplantation and in  
 CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,  
 CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,  
 CC in the treatment of burns, incisions and ulcers; as well as in treatment

CC of periodontal disease, osteoporosis or osteoarthritis, mediated by  
 CC inflammatory processes, diseases of the peripheral nervous system,  
 CC Alzheimer's, Parkinson's disease, Huntington's disease,  
 CC amyotrophic lateral sclerosis, and Shy-Drager syndrome. Infections,  
 CC infarction of cardiac and central nervous system vessel e.g. stroke,  
 CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The  
 CC protein, having activin- or inhibin-related activities is useful as a  
 CC contraceptive based on the ability of inhibin to decrease fertility in  
 CC female mammals and decrease spermatogenesis in male mammals. The  
 CC proteins and nucleic acids are also useful as food supplements. The  
 CC present sequence represents a secreted protein of the invention.  
 XX SQ Sequence 186 AA;  
 Query Match 2.0%; Score 8; DB 22; Length 186;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
 Db 39 SSGASTAT 46

RESULT 28  
 ABB55712  
 ID ABB55712 standard; Protein; 186 AA.  
 XX AC ABB55712;  
 XX DT 14-FEB-2002 (first entry)  
 XX DE Human polypeptide SEQ ID NO 30.  
 XX KW Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;  
 KW immune disorder; bacterial infection; fungal infection; cancer; tumour;  
 KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;  
 KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
 KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulneryary;  
 KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;  
 KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;  
 KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory.  
 XX OS Homo sapiens.  
 XX KW US2001039335-A1.  
 XX PD 08-NOV-2001.  
 XX PF 04-DEC-2000; 2000US-0729674.  
 XX PR 10-APR-1997; 97US-126425P.  
 XX PR 04-DEC-1997; 97US-067454P.  
 XX PR 20-DEC-1997; 97US-068379P.  
 XX PR 02-JAN-1998; 98US-070346P.  
 XX PR 07-JAN-1998; 98US-070643P.  
 XX PR 08-JAN-1998; 98US-070755P.  
 XX PR 13-JAN-1998; 98US-071304P.  
 XX PR 22-JAN-1998; 98US-072134P.  
 XX PR 30-JAN-1998; 98US-073095P.  
 XX PR 18-FEB-1998; 98US-075038P.  
 XX PR 30-MAR-2000; 2000US-0539330.  
 XX PR 23-NOV-1998; 98US-0197886.  
 XX (JACO) JACOBS K.  
 PA (MCCO) MCCOY J M.  
 PA (LAVA) LAVALLIE E R.  
 PA (COLL) COLLINS-RACIE L A.  
 PA (EVAN) EVANS C.  
 PA (MERB) MERBERG D.  
 PA (TREA) TREACY M.  
 PA (AGOS) AGOSTINO M J.  
 PA (STEI) STEININGER R J.

PA (SPAU/) SPAULDING V.  
PA (WONG/) WONG G G.  
PA (CLAR/) CLARK H.  
PA (FECH/) FECHTEL K.  
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
PI Wong GG, Clark H, Fechtel K;  
XX WPI: 2002-040725/05.  
DR N-PSDB; ABA90890.  
XX  
PT New secreted proteins and encoding polynucleotides, useful in gene  
PT therapies, particularly for preventing or treating autoimmune  
PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,  
PT stroke or inflammations -  
XX  
PS Disclosure; Page 197-198; 349pp; English.  
XX  
CC The invention relates to isolated polynucleotides (ABA90876-ABA90968 and  
CC ABA90980) and encoded proteins (ABBS5598-ABBS5800), especially  
CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and  
CC proteins SEQ ID NO 2 (ABBS5598) and SEQ ID NO 20 (ABBS5707) contained in  
CC clones bd306-7 and y8-1 respectively and the clones bd306-7 and y8-1  
CC are deposited with the American Type Culture Collection (ATCC) with  
CC accession number 98599. The polynucleotides and encoded polypeptides have  
CC cytotatic, anti-inflammatory, immunomodulator, vulnerary,  
CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic  
CC and anti-inflammatory activity and acting as cytokine modulators,  
CC haematopoiesis regulators, tissue growth modulators and/or cadherin  
CC suppressors. The polypeptides and polynucleotides are useful in gene  
CC therapies, particularly for preventing, treating or ameliorating any of  
CC the following diseases: immune deficiency and disorders; e.g. bacterial  
CC or fungal infections, autoimmune disorders, cancer, systemic lupus  
CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
CC osteoarthritis; central and peripheral nervous system diseases and  
CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis  
CC or systemic inflammatory response syndrome, ischaemia-reperfusion  
CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or  
CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus  
CC foliaceus.  
XX  
SQ Sequence 186 AA;  
  
Query Match 2.0%; Score 8; DB 23; Length 186;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46  
  
RESULT 29  
ABB71446  
ID ABB71446 standard; Protein; 217 AA.  
XX  
AC ABB71446;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 41130.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX

PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL15549.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 41130; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 217 AA;  
  
Query Match 2.0%; Score 8; DB 22; Length 217;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 341 LQQLAAAV 348  
Db 33 LQQLAAAV 40  
  
RESULT 30  
AAU58834  
ID AAU58834 standard; Protein; 242 AA.  
XX  
AC AAU58834;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #19730.  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX	Hasselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI	Yamamoto RT, Xu HH;
XX	
DR	WPI; 2001-611495/70.
DR	N-PSDB; AAS556165.
XX	
PT	New polynucleotides for the identification and development of
PI	antibiotics, comprise sequences of antisense nucleic acids -
XX	
XX	Example 3; Seq ID No 13899; 51lpp: English.
XX	
CC	The invention relates to antisense inhibitors of genes essential to
CC	prokaryotic cellular proliferation, their use in identifying the
CC	genes, their use in the discovery of novel antibiotics, the essential
CC	genes themselves and the encoded proteins. The prokaryotes used are
CC	Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC	pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC	invention is also useful for the identification of potential new targets
CC	for antibiotic development. The antisense nucleic acids can also be used
CC	to identify proteins used in proliferation, to express these proteins,
CC	and to obtain antibodies capable of binding to the expressed proteins.
CC	The proteins can be used to screen compounds in rational drug discovery
CC	programmes. The antisense nucleic acid sequence is also useful to screen
CC	for homologous nucleic acids which are required for cell proliferation in
CC	a wide variety of organisms. The present sequence represents an
CC	essential prokaryotic cellular proliferation protein.
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 254 AA;
S	

```

Query Match      2.0%; Score 8; DB 22; Length 254;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      359 REA1ORLE 166
Db      52 REA1ORLE 59
      |||||
      |||||

RESULT 32
AAR74043
ID AAR74043 standard; Protein; 272 AA.
XX
XX AAR74043;
XX AC
XX AC
XX AC
DT 25-MAR-2003 (updated)
DT 01-DEC-1995 (first entry)
XX
XX Rice carbonic-anhydrase.
XX
XX Rice; carbonic-anhydrase; carbon-fixation; monocotyledon; cereal;
XX crop improvement.
XX
XX Oryza sativa.
XX
XX WO95111979-A1.
XX
XX 04-MAY-1995.
XX
XX 27-OCT-1994; 94WO-JP01814.
XX
XX 29-OCT-1993; 93JP-0294278.
XX
XX (NISB ) JAPAN TOBACCO INC.
XX
XX Burnell NJ, Suzuki S;
XX
XX WPI; 1995-178871/23.
XX N-PSDB; AAO92067.
XX

```

XX Cloned DNA coding for monocotyledon carbonic anhydrase - is used  
PT to improve carbon fixation in monocotyledonous plants.  
PT  
XX  
XX Claim 8; Page 17-18; 42pp; Japanese.  
XX  
CC The cDNA sequence given in AAQ92067 was obtd. by screening and  
CC cloning a cDNA library generated from rice RNA, and encodes  
CC carbonic-anhydrase.  
CC (Updated on 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 272 AA;  
Query Match 2.0%; Score 8; DB 16; Length 272;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 100 AAPVAPAA 107  
Db 63 AAPVAPAA 70  
|||||  
RESULT 33  
AAG20687  
ID AAG20687 standard; Protein; 309 AA.  
XX  
AC AAG20687;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22973.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0145386.  
PR 02-AUG-1999; 99US-0145388.  
PR 02-AUG-1999; 99US-0145389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 2.0%; Score 8; DB 21; Length 309;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 7 TLKGTNFE 14  
|||||||  
Db 295 TLKGTNFE 302  
  
RESULT 34  
AAG38007  
ID AAG38007 standard; Protein; 309 AA.  
XX AC AAG38007;  
XX 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 46824.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 09-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 23-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.



PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149910.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

2.0%; Score 8; DB 21; Length 309;

Best Local Similarity 100.0%; Pred. No. 85;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 TLKGTNFE 14  
Db 295 TLKGTNFE 302

RESULT 35  
AAG20686  
ID AAG20686 standard; Protein; 316 AA.  
AC AAG20686;  
XX  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22972.  
XX  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX  
PN EP1033405-A2.  
XX  
XX  
PD 06-SEP-2000.  
PF  
PF  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132563.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.

```
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match      2.0%; Score 8; DB 21; Length 316;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 TLKGTNFE 14
Db 302 TLKGTNFE 309
```

## RESULT 36

AAG38006

ID AAG38006 standard; Protein; 316 AA.

XX

AC AAG38006;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 46823.

XX

KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

KW

XX Arabidopsis thaliana.

OS

XX Arabidopsis thaliana.

PN

EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR

PR 05-MAR-1999; 99US-0123180.

PR

PR 09-MAR-1999; 99US-0123548.

PR

PR 23-MAR-1999; 99US-0125788.

PR

PR 25-MAR-1999; 99US-0126264.

PR

PR 29-MAR-1999; 99US-0126785.

PR

PR 01-APR-1999; 99US-0127462.

PR

PR 08-APR-1999; 99US-0128714.

PR

PR 16-APR-1999; 99US-0129845.

PR

PR 19-APR-1999; 99US-0130077.

PR

PR 21-APR-1999; 99US-0130449.

PR

PR 23-APR-1999; 99US-0130510.

PR

PR 28-APR-1999; 99US-0130891.

PR

PR 30-APR-1999; 99US-0131449.

PR

PR 30-APR-1999; 99US-0132048.

PR

PR 04-MAY-1999; 99US-0132407.

PR

PR 05-MAY-1999; 99US-0132484.

PR

PR 06-MAY-1999; 99US-0132485.

PR

PR 06-MAY-1999; 99US-0132486.

PR

PR 07-MAY-1999; 99US-0132487.

PR

PR 11-MAY-1999; 99US-0132863.

PR

PR 14-MAY-1999; 99US-0134218.

PR

PR 14-MAY-1999; 99US-0134219.

PR

PR 14-MAY-1999; 99US-0134221.

PR

PR 18-MAY-1999; 99US-0134370.

PR

PR 19-MAY-1999; 99US-0134768.

PR

PR 20-MAY-1999; 99US-0134941.

PR

PR 21-MAY-1999; 99US-0135124.

PR

PR 21-MAY-1999; 99US-0135353.

PR

PR 24-MAY-1999; 99US-0135629.

PR

PR 25-MAY-1999; 99US-0136021.

PR

PR 27-MAY-1999; 99US-0136392.

PR

PR 28-MAY-1999; 99US-0136782.

PR

PR 01-JUN-1999; 99US-0137222.

PR

PR 03-JUN-1999; 99US-0137528.

PR

PR 04-JUN-1999; 99US-0137502.

PR

PR 07-JUN-1999; 99US-0137724.

PR

PR 08-JUN-1999; 99US-0138094.

PR

PR 10-JUN-1999; 99US-0138540.

PR

PR 10-JUN-1999; 99US-0138847.

PR

PR 14-JUN-1999; 99US-0139119.

PR

PR 16-JUN-1999; 99US-0139452.

PR

PR 16-JUN-1999; 99US-0139453.

PR

PR 17-JUN-1999; 99US-0139492.

PR

PR 18-JUN-1999; 99US-0139454.

PR

PR 18-JUN-1999; 99US-0139455.

PR

PR 18-JUN-1999; 99US-0139456.

PR

PR 18-JUN-1999; 99US-0139457.

PR

PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 8; DB 21; Length 316;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 TLKGTNFE 14  
| | | | |  
Db 302 TLKGTNFE 309

## RESULT 37

AAG18084  
ID AAG18084 standard; Protein; 338 AA.  
XX  
AC AAG18084;  
XX



PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159322.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 8; DB 21; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 LSKAKASS 81  
 |||||  
 Db 11 LSKAKASS 18

RESULT 38  
 AAG20685  
 ID AAG20685 standard; Protein; 342 AA.

XX AAG20685;

XX DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 22971.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX OS Arabidopsis thaliana.  
 XX FN EP1033405-A2.  
 XX PD 06-SEP-2000.  
 XX PP 25-FEB-2000; 2000EP-0301439.  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139452.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.

```
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
```

```
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 2.0%; Score 8; DB 21; Length 342;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TLKGTNFE 14  
|||  
Db 328 TLKGTNFE 335

## RESULT 39

AAG38005  
ID AAG38005 standard; Protein; 342 AA.

XX AC AAG38005;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 46822.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX XX

PF 25-FEB-2000; 2000EP-0301439.  
XX 99US-0121825.  
PR 05-FEB-1999; 99US-0123180.  
PR 05-MAR-1999; 99US-0123548.  
PR 03-MAR-1999; 99US-0125788.  
PR 23-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 16-MAY-1999; 99US-0134768.  
PR 18-MAY-1999; 99US-0134941.  
PR 19-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.



PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161320.  
PR 28-OCT-1999; 99US-0161392.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162143.

Query Match 2.0%; Score 8; DB 21; Length 342;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TLKGTNFE 14  
|||||||  
DB 328 TLKGTNFE 335

RESULT 40  
ABP80988  
ID ABP80988 standard; Protein; 423 AA.  
XX AC ABP80988;  
XX DT 07-MAR-2003 (first entry)  
XX DE N. gonorrhoeae amino acid sequence SEQ ID 8506.  
XX KW Antibacterial; infection; vaccine; gene therapy.  
XX OS Neisseria gonorrhoeae.  
XX PN W0200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB02069.  
XX PR 12-FEB-2001; 2001GB-0003424.  
XX PA (CHIR-) CHIRON SPA.  
XX PI Fontana MR, Pizza M, Masighani V, Monaci E;  
XX WPI; 2003-058415/05.

DR N-PSDB; ABZ41958.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
PT medicament for treating or preventing N. gonorrhoeae infection -  
XX Disclosure; Page 807; 815pp; English.  
XX The present invention relates to proteins from Neisseria gonorrhoeae.  
CC Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
CC infection, this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
CC molecules of the invention.

SQ Sequence 423 AA;

Query Match 2.0%; Score 8; DB 24; Length 423;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 QLAAAVPQ 350  
|||||||  
DB 28 QLAAAVPQ 35

RESULT 41

AMG48742

ID AAG48742 standard; Protein; 536 AA.

XX AC AAG48742;  
XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 61585.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.  
XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 25-MAR-1999; 99US-0126264.  
XX PR 29-MAR-1999; 99US-0126785.  
XX PR 01-APR-1999; 99US-0127462.  
XX PR 06-APR-1999; 99US-0128234.  
XX PR 08-APR-1999; 99US-0128714.  
XX PR 16-APR-1999; 99US-0129845.  
XX PR 19-APR-1999; 99US-0130077.  
XX PR 21-APR-1999; 99US-0130449.  
XX PR 23-APR-1999; 99US-0130510.  
XX PR 23-APR-1999; 99US-0130891.  
XX PR 28-APR-1999; 99US-0131449.  
XX PR 30-APR-1999; 99US-0132048.  
XX PR 30-APR-1999; 99US-0132407.  
XX PR 04-MAY-1999; 99US-0132484.  
XX PR 05-MAY-1999; 99US-0132485.  
XX PR 06-MAY-1999; 99US-0132486.  
XX PR 06-MAY-1999; 99US-0132487.  
XX PR 07-MAY-1999; 99US-0132863.  
XX PR 11-MAY-1999; 99US-0134256.  
XX PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219;  
PR 14-MAY-1999; 99US-0134221;  
PR 14-MAY-1999; 99US-0134370;  
PR 18-MAY-1999; 99US-0134768;  
PR 19-MAY-1999; 99US-0134941;  
PR 20-MAY-1999; 99US-0135124;  
PR 21-MAY-1999; 99US-0135353;  
PR 24-MAY-1999; 99US-0135629;  
PR 25-MAY-1999; 99US-0136021;  
PR 27-MAY-1999; 99US-0136392;  
PR 28-MAY-1999; 99US-0136782;  
PR 01-JUN-1999; 99US-0137222;  
PR 03-JUN-1999; 99US-0137528;  
PR 04-JUN-1999; 99US-0137502;  
PR 07-JUN-1999; 99US-0137724;  
PR 08-JUN-1999; 99US-0138094;  
PR 10-JUN-1999; 99US-0138540;  
PR 10-JUN-1999; 99US-0138847;  
PR 14-JUN-1999; 99US-0139119;  
PR 16-JUN-1999; 99US-0139452;  
PR 16-JUN-1999; 99US-0139453;  
PR 17-JUN-1999; 99US-0139492;  
PR 18-JUN-1999; 99US-0139454;  
PR 18-JUN-1999; 99US-0139455;  
PR 18-JUN-1999; 99US-0139456;  
PR 18-JUN-1999; 99US-0139457;  
PR 18-JUN-1999; 99US-0139458;  
PR 18-JUN-1999; 99US-0139459;  
PR 18-JUN-1999; 99US-0139460;  
PR 18-JUN-1999; 99US-0139461;  
PR 18-JUN-1999; 99US-0139462;  
PR 18-JUN-1999; 99US-0139463;  
PR 18-JUN-1999; 99US-0139750;  
PR 18-JUN-1999; 99US-0139763;  
PR 21-JUN-1999; 99US-0139817;  
PR 22-JUN-1999; 99US-0139817;  
PR 22-JUN-1999; 99US-0139899;  
PR 23-JUN-1999; 99US-0140355;  
PR 23-JUN-1999; 99US-0140354;  
PR 24-JUN-1999; 99US-0140695;  
PR 28-JUN-1999; 99US-0140823;  
PR 29-JUN-1999; 99US-0140991;  
PR 30-JUN-1999; 99US-0141287;  
PR 01-JUL-1999; 99US-0141842;  
PR 01-JUL-1999; 99US-0142154;  
PR 02-JUL-1999; 99US-0142055;  
PR 06-JUL-1999; 99US-0142390;  
PR 08-JUL-1999; 99US-0142803;  
PR 09-JUL-1999; 99US-0142920;  
PR 12-JUL-1999; 99US-0142977;  
PR 13-JUL-1999; 99US-0143542;  
PR 14-JUL-1999; 99US-0143624;  
PR 15-JUL-1999; 99US-0144005;  
PR 16-JUL-1999; 99US-0144005;  
PR 16-JUL-1999; 99US-0144085;  
PR 16-JUL-1999; 99US-0144086;  
PR 19-JUL-1999; 99US-0144325;  
PR 19-JUL-1999; 99US-0144331;  
PR 19-JUL-1999; 99US-0144332;  
PR 19-JUL-1999; 99US-0144333;  
PR 19-JUL-1999; 99US-0144334;  
PR 19-JUL-1999; 99US-0144335;  
PR 20-JUL-1999; 99US-0144352;  
PR 20-JUL-1999; 99US-0144632;  
PR 20-JUL-1999; 99US-0144884;  
PR 21-JUL-1999; 99US-0144814;  
PR 21-JUL-1999; 99US-0145086;  
PR 21-JUL-1999; 99US-0145088;  
PR 22-JUL-1999; 99US-0145085;  
PR 22-JUL-1999; 99US-0145087;  
PR 22-JUL-1999; 99US-0145089;  
PR 23-JUL-1999; 99US-0145192;  
PR 23-JUL-1999; 99US-0145145;  
PR 23-JUL-1999; 99US-0145210;  
PR 23-JUL-1999; 99US-0145224;  
PR 26-JUL-1999; 99US-0145276;  
PR 27-JUL-1999; 99US-0145913;  
PR 27-JUL-1999; 99US-0145918;  
PR 27-JUL-1999; 99US-0145919;  
PR 28-JUL-1999; 99US-0145951;  
PR 02-AUG-1999; 99US-0146386;  
PR 02-AUG-1999; 99US-0146388;  
PR 02-AUG-1999; 99US-0146389;  
PR 03-AUG-1999; 99US-0147038;  
PR 04-AUG-1999; 99US-0147204;  
PR 04-AUG-1999; 99US-0147302;  
PR 05-AUG-1999; 99US-0147192;  
PR 05-AUG-1999; 99US-0147260;  
PR 06-AUG-1999; 99US-0147303;  
PR 06-AUG-1999; 99US-0147416;  
PR 09-AUG-1999; 99US-0147493;  
PR 09-AUG-1999; 99US-0147935;  
PR 10-AUG-1999; 99US-0148171;  
PR 11-AUG-1999; 99US-0148319;  
PR 12-AUG-1999; 99US-0148341;  
PR 13-AUG-1999; 99US-0148565;  
PR 13-AUG-1999; 99US-0148684;  
PR 16-AUG-1999; 99US-0149368;  
PR 17-AUG-1999; 99US-0149175;  
PR 18-AUG-1999; 99US-0149426;  
PR 20-AUG-1999; 99US-0149722;  
PR 20-AUG-1999; 99US-0149723;  
PR 20-AUG-1999; 99US-0149929;  
PR 23-AUG-1999; 99US-0149902;  
PR 23-AUG-1999; 99US-0149930;  
PR 25-AUG-1999; 99US-0150566;  
PR 26-AUG-1999; 99US-0150884;  
PR 27-AUG-1999; 99US-0151065;  
PR 27-AUG-1999; 99US-0151066;  
PR 27-AUG-1999; 99US-0151080;  
PR 30-AUG-1999; 99US-0151303;  
PR 31-AUG-1999; 99US-0151438;  
PR 01-SEP-1999; 99US-0151930;  
PR 07-SEP-1999; 99US-0152363;  
PR 10-SEP-1999; 99US-0153070;  
PR 13-SEP-1999; 99US-0153758;  
PR 15-SEP-1999; 99US-0154018;  
PR 16-SEP-1999; 99US-0154039;  
PR 20-SEP-1999; 99US-0154779;  
PR 22-SEP-1999; 99US-0155139;  
PR 23-SEP-1999; 99US-0155486;  
PR 24-SEP-1999; 99US-0155659;  
PR 28-SEP-1999; 99US-0156458;  
PR 29-SEP-1999; 99US-0156596;  
PR 04-OCT-1999; 99US-0157117;  
PR 05-OCT-1999; 99US-0157753;  
PR 06-OCT-1999; 99US-0157865;  
PR 07-OCT-1999; 99US-0158029;  
PR 08-OCT-1999; 99US-0158232;  
PR 12-OCT-1999; 99US-0158369;  
PR 13-OCT-1999; 99US-0159293;  
PR 13-OCT-1999; 99US-0159294;  
PR 13-OCT-1999; 99US-0159295;  
PR 14-OCT-1999; 99US-0159329;  
PR 14-OCT-1999; 99US-0159330;  
PR 14-OCT-1999; 99US-0159331;  
PR 14-OCT-1999; 99US-0159637;  
PR 18-OCT-1999; 99US-0159638;  
PR 18-OCT-1999; 99US-0159584;  
PR 21-OCT-1999; 99US-0160741;  
PR 21-OCT-1999; 99US-0160767;  
PR 21-OCT-1999; 99US-0160768;  
PR 21-OCT-1999; 99US-0160770;  
PR 21-OCT-1999; 99US-0160814;  
PR 21-OCT-1999; 99US-0160815;  
PR 22-OCT-1999; 99US-0160980;  
PR 22-OCT-1999; 99US-0160981;  
PR 22-OCT-1999; 99US-0160989;  
PR 22-OCT-1999; 99US-0160989;

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161362.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 8; DB 21; Length 536;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 LSKAKASS 81  
Db 10 LSKAKASS 17  
|||||

RESULT 42  
AAW89794  
ID AAW89794 standard; Protein; 540 AA.  
XX  
AC AAW89794;  
XX  
DT 16-MAR-1999 (first entry)  
XX  
DE Staphylococcus aureus protein SEQ ID #5242.

XX Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome.  
XX  
OS Staphylococcus aureus.  
XX  
PN EP786519-A2.  
XX  
PD 30-JUL-1997.

XX 07-JAN-1997; 97EP-0100117.  
XX  
XX 05-JAN-1996; 96US-0009861.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
PI  
PI Rosen CA;  
XX  
XX WPI; 1997-374922/35.  
XX

XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
XX  
XX Claim 23; Page 3244-3246; 3271pp; English.  
XX  
XX This sequence represents a Staphylococcus aureus protein sequence of the  
CC invention. The DNA sequences encoding the S.aureus proteins are recorded  
CC on a computer readable medium, preferably selected from a floppy or hard  
CC disk, random access memory (RAM), read-only memory (ROM) or CD-ROM.  
CC Homology searches using the S.aureus DNA sequences allows putative  
CC functions to be assigned so that protein-encoding or regulatory regions  
CC of commercial, therapeutic or industrial importance can be obtained.  
CC Specifically, sequences which are likely to encode antigens have been  
CC identified and these polypeptides can be used in a vaccine composition  
CC against S.aureus infection. The polypeptides can also be used in a kit  
CC for the immunodetection of S.aureus in a sample. S.aureus is implicated  
CC in numerous human diseases, including cellulitis, eyelid infections, food  
CC poisoning, osteomyelitis, skin and surgical wound infections, scalded  
CC skin syndrome, toxic shock syndrome, etc. Organisms transformed with the  
CC DNA sequences can be used for recombinant production of the polypeptides.

CC The new DNA sequences (and their fragments) are useful as primers or  
CC probes for isolating homologues of any of the 5191 S.aureus DNA sequences  
CC contained on the computer readable medium.

XX Sequence 540 AA;

QY Query Match 2.0%; Score 8; DB 18; Length 540;  
Db Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQAAPVA 104  
Db 175 AQAAPVA 182  
|||||

RESULT 43  
AAW89777  
ID AAW89777 standard; Protein; 540 AA.  
XX  
AC AAW89777;  
XX  
DT 16-MAR-1999 (first entry)  
XX  
DE Staphylococcus aureus protein SEQ ID #5225.

XX Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome.  
XX  
OS Staphylococcus aureus.  
XX  
PN EP786519-A2.  
XX  
PD 30-JUL-1997.

XX 07-JAN-1997; 97EP-0100117.  
XX  
XX 05-JAN-1996; 96US-0009861.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
PI  
PI Rosen CA;  
XX  
XX WPI; 1997-374922/35.

XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
PT stored on computer readable medium and used in the production of  
PT anti-S.aureus vaccines  
XX  
XX Claim 23; Page 3221-3222; 3271pp; English.  
XX  
XX This sequence represents a Staphylococcus aureus protein sequence of the  
CC invention. The DNA sequences encoding the S.aureus proteins are recorded  
CC on a computer readable medium, preferably selected from a floppy or hard  
CC disk, random access memory (RAM), read-only memory (ROM) or CD-ROM.  
CC Homology searches using the S.aureus DNA sequences allows putative  
CC functions to be assigned so that protein-encoding or regulatory regions

/note= "these residues represent a line of missing text  
in the sequence listing in the specification.  
They are included to maintain the residue  
numbering given in the specification for this  
protein sequence"

/note= "these residues represent a line of missing text  
in the sequence listing in the specification.  
They are included to maintain the residue  
numbering given in the specification for this  
protein sequence"

CC of commercial, therapeutic or industrial importance can be obtained.  
 CC Specifically, sequences which are likely to encode antigens have been  
 CC identified and these polypeptides can be used in a vaccine composition  
 CC against *S. aureus* infection. The polypeptides can also be used in a kit  
 CC for the immunodetection of *S. aureus* in a sample. *S. aureus* is implicated  
 CC in numerous human diseases, including cellulitis, eyelid infections, food  
 CC poisoning, osteomyelitis, skin and surgical wound infections, scalded  
 CC skin syndrome, toxic shock syndrome, etc. Organisms transformed with the  
 CC DNA sequences can be used for recombinant production of the polypeptides.  
 CC The new DNA sequences (and their fragments) are useful as primers or  
 CC probes for isolating homologues of any of the 5191 *S. aureus* DNA sequences  
 CC contained on the computer readable medium.

SQ Sequence 540 AA;

Query Match 2.0%; Score 8; DB 18; Length 540;

Best Local Similarity 100.0%; Pred. No. 1.5e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQPAPVA 104

|||||

Db 175 AQPAPVA 182

RESULT 44

AA99408

ID AA99408 standard; Protein; 596 AA.

XX

AC AA99408;

XX

DT 08-AUG-2000 (first entry)

XX

DE Human PRO1342 (UNQ697) amino acid sequence SEQ ID NO:243.

XX

KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;  
 KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.

XX

OS Homo sapiens.

PN

WO200012708-A2.

XX

PD 09-MAR-2000.

XX

PF 01-SEP-1999; 99WO-US20111.

XX

PR 01-SEP-1998; 98US-0098716.

PR

PR 01-SEP-1998; 98US-0098749.

PR

PR 01-SEP-1998; 98US-0098750.

PR

PR 02-SEP-1998; 98US-0098803.

PR

PR 02-SEP-1998; 98US-0098821.

PR

PR 02-SEP-1998; 98US-0098843.

PR

PR 02-SEP-1998; 98US-0098536.

PR

PR 09-SEP-1998; 98US-0099596.

PR

PR 09-SEP-1998; 98US-0099598.

PR

PR 09-SEP-1998; 98US-0099602.

PR

PR 10-SEP-1998; 98US-0099642.

PR

PR 10-SEP-1998; 98US-0099741.

PR

PR 10-SEP-1998; 98US-0099754.

PR

PR 10-SEP-1998; 98US-0099763.

PR

PR 10-SEP-1998; 98US-0099792.

PR

PR 10-SEP-1998; 98US-0099808.

PR

PR 10-SEP-1998; 98US-0099812.

PR 17-SEP-1998; 98US-0100684.  
 PR 17-SEP-1998; 98US-0100710.  
 PR 17-SEP-1998; 98US-0100711.  
 PR 17-SEP-1998; 98US-0100919.  
 PR 17-SEP-1998; 98US-0100930.  
 PR 18-SEP-1998; 98US-0100848.  
 PR 18-SEP-1998; 98US-0100849.  
 PR 18-SEP-1998; 98US-0101014.  
 PR 18-SEP-1998; 98US-0101068.  
 PR 18-SEP-1998; 98US-0101071.  
 PR 18-SEP-1998; 98US-0101279.  
 PR 23-SEP-1998; 98US-0101471.  
 PR 23-SEP-1998; 98US-0101472.  
 PR 23-SEP-1998; 98US-0101474.  
 PR 23-SEP-1998; 98US-0101475.  
 PR 23-SEP-1998; 98US-0101476.  
 PR 23-SEP-1998; 98US-0101477.  
 PR 23-SEP-1998; 98US-0101479.  
 PR 24-SEP-1998; 98US-0101738.  
 PR 24-SEP-1998; 98US-0101741.  
 PR 24-SEP-1998; 98US-0101743.  
 PR 24-SEP-1998; 98US-0101915.  
 PR 24-SEP-1998; 98US-0101916.  
 PR 29-SEP-1998; 98US-0102207.  
 PR 29-SEP-1998; 98US-0102240.  
 PR 29-SEP-1998; 98US-0102307.  
 PR 29-SEP-1998; 98US-0102330.  
 PR 30-SEP-1998; 98US-0102331.  
 PR 30-SEP-1998; 98US-0102484.  
 PR 30-SEP-1998; 98US-0102487.  
 PR 30-SEP-1998; 98US-0102570.  
 PR 01-OCT-1998; 98US-0102571.  
 PR 01-OCT-1998; 98US-0102684.  
 PR 01-OCT-1998; 98US-0102687.  
 PR 02-OCT-1998; 98US-0102965.  
 PR 06-OCT-1998; 98US-0103258.  
 PR 06-OCT-1998; 98US-0103449.  
 PR 07-OCT-1998; 98US-0103314.  
 PR 07-OCT-1998; 98US-0103315.  
 PR 07-OCT-1998; 98US-0103328.  
 PR 07-OCT-1998; 98US-0103395.  
 PR 07-OCT-1998; 98US-0103396.  
 PR 07-OCT-1998; 98US-0103398.  
 PR 08-OCT-1998; 98US-0103633.  
 PR 08-OCT-1998; 98US-0103678.  
 PR 08-OCT-1998; 98US-0103679.  
 PR 08-OCT-1998; 98US-0103711.  
 PR 20-OCT-1998; 98US-0104257.  
 PR 20-OCT-1998; 98US-0104987.  
 PR 20-OCT-1998; 98US-0105000.  
 PR 20-OCT-1998; 98US-0105002.  
 PR 21-OCT-1998; 98US-0105104.  
 PR 22-OCT-1998; 98US-0105169.  
 PR 22-OCT-1998; 98US-0105266.  
 PR 22-OCT-1998; 98US-0105693.  
 PR 26-OCT-1998; 98US-0105694.  
 PR 27-OCT-1998; 98US-0105807.  
 PR 27-OCT-1998; 98US-0105881.  
 PR 27-OCT-1998; 98US-0105882.  
 PR 28-OCT-1998; 98US-0106023.  
 PR 28-OCT-1998; 98US-0106029.  
 PR 28-OCT-1998; 98US-0106030.  
 PR 28-OCT-1998; 98US-0106032.  
 PR 28-OCT-1998; 98US-0106033.  
 PR 29-OCT-1998; 98US-0106178.  
 PR 29-OCT-1998; 98US-0106248.  
 PR 29-OCT-1998; 98US-0106384.  
 PR 30-OCT-1998; 98US-0108500.  
 PR 03-NOV-1998; 98US-0106464.  
 PR 03-NOV-1998; 98US-0106855.  
 PR 03-NOV-1998; 98US-0106902.  
 PR 03-NOV-1998; 98US-0106905.

PR 03-NOV-1998; 98US-0106919.  
PR 03-NOV-1998; 98US-0106932.  
PR 03-NOV-1998; 98US-0106934.  
PR 10-NOV-1998; 98US-0107783.  
PR 17-NOV-1998; 98US-0108775.  
PR 17-NOV-1998; 98US-0108779.  
PR 17-NOV-1998; 98US-0108787.  
PR 17-NOV-1998; 98US-0108788.  
PR 17-NOV-1998; 98US-0108801.  
PR 17-NOV-1998; 98US-0108802.  
PR 17-NOV-1998; 98US-0108806.  
PR 17-NOV-1998; 98US-0108807.  
PR 17-NOV-1998; 98US-0108867.  
PR 17-NOV-1998; 98US-0108925.  
PR 18-NOV-1998; 98US-0108848.  
PR 18-NOV-1998; 98US-0108849.  
PR 18-NOV-1998; 98US-0108850.  
PR 18-NOV-1998; 98US-0108851.  
PR 18-NOV-1998; 98US-0108852.  
PR 18-NOV-1998; 98US-0108858.  
PR 18-NOV-1998; 98US-0108904.  
XX  
PA (GETH ) GENENTECH INC.  
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;  
XX WPI; 2000-237871/20.  
XX N-PSDB; AAA37090.  
XX  
XX New mammalian DNA sequences encoding transmembrane, receptor or  
PT secreted PRO polypeptides, useful for screening of potential peptide or  
PT small molecule inhibitors of the relevant receptor/ligand interactions  
XX  
XX Claim 12; Fig 138; 773pp; English.  
XX  
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,  
CC receptor or secreted PRO polypeptides given in AA99340 to AA99462. The  
CC transmembrane and receptor PRO proteins can be used for screening of  
CC potential peptide or small molecule inhibitors of the relevant  
CC receptor/ligand interactions. The polypeptides and nucleotide sequences  
CC encoding then have various industrial applications, including uses as  
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent  
CC PCR primers and hybridisation probes used in the isolation of the PRO  
CC polypeptides from the present invention.  
XX  
XX Sequence 596 AA;  
SQ  
Query Match 2.0%; Score 8; DB 21; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46  
|||||  
RESULT 45  
AAU29178  
ID AAU29178 standard; Protein; 596 AA.  
XX  
XX AAU29178;  
AC  
XX  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human PRO polypeptide sequence #155.  
XX  
XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;  
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;  
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.  
XX  
OS Homo sapiens.  
XX

PN WO200168849-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 28-FEB-2001; 2001WO-US06520.  
XX  
PR 01-MAR-2000; 2000WO-US05601.  
PR 02-MAR-2000; 2000WO-US05841.  
PR 03-MAR-2000; 2000US-187202P.  
PR 06-MAR-2000; 2000US-186968P.  
PR 14-MAR-2000; 2000US-189320P.  
PR 14-MAR-2000; 2000US-189328P.  
PR 15-MAR-2000; 2000WO-US06884.  
PR 21-MAR-2000; 2000US-190828P.  
PR 21-MAR-2000; 2000US-191007P.  
PR 21-MAR-2000; 2000US-191048P.  
PR 21-MAR-2000; 2000US-191314P.  
PR 28-MAR-2000; 2000US-192855P.  
PR 29-MAR-2000; 2000US-193032P.  
PR 29-MAR-2000; 2000US-193053P.  
PR 30-MAR-2000; 2000WO-US08439.  
PR 04-APR-2000; 2000US-194449P.  
PR 04-APR-2000; 2000US-194647P.  
PR 11-APR-2000; 2000US-195975P.  
PR 11-APR-2000; 2000US-196000P.  
PR 11-APR-2000; 2000US-196187P.  
PR 11-APR-2000; 2000US-196690P.  
PR 11-APR-2000; 2000US-196820P.  
PR 18-APR-2000; 2000US-198121P.  
PR 18-APR-2000; 2000US-198585P.  
PR 25-APR-2000; 2000US-199397P.  
PR 25-APR-2000; 2000US-199550P.  
PR 25-APR-2000; 2000US-199654P.  
PR 03-MAY-2000; 2000US-201516P.  
PR 17-MAY-2000; 2000WO-US13705.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 30-MAY-2000; 2000WO-US14941.  
PR 02-JUN-2000; 2000WO-US15264.  
PR 05-JUN-2000; 2000US-209832P.  
PR 28-JUL-2000; 2000WO-US20710.  
PR 22-AUG-2000; 2000US-0644848.  
PR 24-AUG-2000; 2000WO-US23328.  
PR 08-NOV-2000; 2000WO-US30952.  
PR 01-DEC-2000; 2000WO-US32678.  
PR 20-DEC-2000; 2000WO-US34956.  
XX  
XX (GETH ) GENENTECH INC.  
PA  
XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
XX  
XX WPI; 2001-602746/68.  
DR N-PSDB; AAS46079.  
XX  
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
PT presence of tumours, such as prostate and breast tumours, in mammals and  
PT to screen for modulators of the compounds -  
XX  
XX Claim 11; Fig 310; 774pp; English.  
XX  
XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.  
CC The PRO polypeptides and their associated nucleic acids can be used to  
CC detect the presence of a tumour in a mammal by comparing the level of  
CC expression of a PRO polypeptide in a test sample of cells from the animal  
CC and a control sample of normal cells, whereby a higher level of  
CC expression in the test sample indicates the presence of a tumour in the  
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats  
CC and rabbits but are preferably human. The polypeptides can be used to  
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,  
CC when contacted with it. A specific polypeptide can be used to stimulate  
CC the proliferation or differentiation of chondrocyte cells. The PRO  
CC proteins can be used to determine the presence of tumours and also  
CC susceptibility to tumour development, particularly adrenal, lung, colon,

CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
 CC can be used for genetic analysis of individuals with genetic disorders.

XX  
 XX Sequence 596 AA;

Query Match 2.0%; Score 8; DB 22; Length 596;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
 DB 39 SSGASTAT 46  
 |||||

RESULT 46  
 AAB87575  
 ID AAB87575 standard; Protein; 596 AA.

XX  
 AC AAB87575;

XX 15-MAY-2001 (first entry)

XX Human PRO1342.

XX Human; PRO protein; mapping.

XX Homo sapiens.

XX WO200116318-A2.

XX 08-MAR-2001.

XX 24-AUG-2000; 2000WO-US23328.

XX 01-SEP-1999; 99WO-US20111.

XX 15-SEP-1999; 99WO-US21090.

XX 07-DEC-1999; 99US-0169495.

XX 09-DEC-1999; 99US-0170262.

XX 11-JAN-2000; 2000US-0175481.

XX 18-FEB-2000; 2000WO-US04341.

XX 22-FEB-2000; 2000WO-US04342.

XX 01-MAR-2000; 2000WO-US05601.

XX 03-MAR-2000; 2000US-0187202.

XX 25-APR-2000; 2000US-0199397.

XX 22-MAY-2000; 2000WO-US14042.

XX 05-JUN-2000; 2000US-0209832.

XX (GETH ) GENENTECH INC.

XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;

XX Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;

XX WPI; 2001-183260/18.

XX N-PSDB; AAF92107.

XX Eighty four nucleic acids encoding PRO polypeptides, useful in

XX molecular biology, including use as hybridization probes, and in

XX chromosome and gene mapping. -

XX Claim 12; Fig 100; 278pp; English.

XX The present sequence is a human PRO polypeptide (secreted and

XX transmembrane). The PRO protein, and PRO agonists, PRO antagonists or

XX anti-PRO antibodies are useful for preparation of a medicament useful in

CC the treatment of a condition which is responsive to the PRO protein,

XX agonists, antagonists or anti-PRO antibodies. The PRO protein may also be

XX employed as molecular weight markers for protein electrophoresis. The PRO

XX coding sequence has applications in molecular biology, including use as

XX hybridisation probes, and in chromosome and gene mapping.

XX Sequence 596 AA;

Query Match

Best Local Similarity 100.0%; Score 8; DB 22; Length 596;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88

DB 39 SSGASTAT 46

|||||

RESULT 47

AAB66157

ID AAB66157 standard; protein; 596 AA.

XX

AC AAB66157;

XX 02-APR-2001 (first entry)

XX Protein of the invention #69.

XX Secreted; transmembrane; gene therapy.

XX Unidentified.

XX WO200078961-A1.

XX 28-DEC-2000.

XX 18-FEB-2000; 2000WO-US04342.

XX 23-JUN-1999; 99US-0141037.

XX 26-JUL-1999; 99US-0144758.

XX 01-SEP-1999; 99WO-US20111.

XX 29-OCT-1999; 99US-0162506.

XX 30-NOV-1999; 99WO-US28313.

XX 16-DEC-1999; 99WO-US28551.

XX 05-JAN-2000; 2000WO-US00219.

XX 06-JAN-2000; 2000WO-US00376.

XX (GETH ) GENENTECH INC.

XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

XX Watanabe CK, Williams PM, Wood WI;

XX WPI; 2001-071395/08.

XX Secreted and transmembrane proteins and nucleic acids designated PRO,

XX useful as hybridization probes, in chromosome and gene mapping and gene

XX therapy -

XX Claim 1; Fig 138; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.

XX These proteins and the DNA encoding them may be used as hybridization

XX probes, in chromosome and gene mapping and in the generation of

XX anti-sense RNA and DNA. They may also be used to generate either

XX transgenic animals or knockout animals which are in turn useful for

XX development and screening of therapeutically useful reagents.

XX The nucleic acids may also be used in gene therapy.

XX Sequence 596 AA;

QY 81 SSGASTAT 88

DB 39 SSGASTAT 46

Query Match

Best Local Similarity 100.0%; Score 8; DB 22; Length 596;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 48

ABG95900  
 ID ABG95900 standard; Protein; 596 AA.  
 XX AC ABG95900;  
 DT 10-DEC-2002 (first entry)  
 XX DE Human secreted/transmembrane protein PRO1342.  
 XX DE Human; secreted protein; transmembrane protein; antirheumatic;  
 KW humanarthritis; osteopathic; sports-related joint problem;  
 KW articular cartilage defect; osteoarthritis; rheumatoid arthritis.  
 XX OS Homo sapiens.  
 XX PN US2002119130-A1.  
 XX PD 29-AUG-2002.  
 XX PF 06-DEC-2001; 2001US-0006867.  
 XX PR 29-OCT-1997; 97US-063435P.  
 PR 29-OCT-1997; 97US-064215P.  
 PR 22-APR-1998; 98US-082797P.  
 PR 29-APR-1998; 98US-083495P.  
 PR 15-MAY-1998; 98US-085579P.  
 PR 10-JUN-1998; 98US-088811P.  
 PR 10-JUN-1998; 98US-088824P.  
 PR 11-JUN-1998; 98US-088863P.  
 PR 12-JUN-1998; 98US-089105P.  
 PR 16-JUN-1998; 98US-089514P.  
 PR 16-SEP-1998; 98WO-US19330.  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 14-MAY-1999; 99WO-US10733.  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 15-SEP-1999; 99WO-US21090.  
 PR 15-SEP-1999; 99WO-US21194.  
 PR 22-DEC-1999; 99WO-US30720.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 22-MAY-2000; 2000WO-US14042.  
 PR 02-JUN-2000; 2000WO-US15264.  
 PR 23-AUG-2000; 2000WO-US23522.  
 PR 24-AUG-2000; 2000WO-US23328.  
 PR 10-NOV-2000; 2000WO-US30873.  
 PR 01-DEC-2000; 2000WO-US32378.  
 PR 20-DEC-2000; 2000WO-US34956.  
 PR 28-FEB-2001; 2001WO-US06520.  
 PR 20-JUN-2001; 2001WO-US19692.  
 PR 29-JUN-2001; 2001WO-US21066.  
 PR 09-JUL-2001; 2001WO-US21735.  
 XX PA (GETH ) GENENTECH INC.  
 XX PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;  
 PI Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;  
 XX WPI: 2002-731348/79.  
 DR N-PSDB; ABS74427.  
 XX New isolated secreted and transmembrane PRO polypeptide useful for  
 PT modulating biological activity of a cell, or for treating  
 PT sports-related joint problems, osteoarthritis or rheumatoid arthritis  
 XX Claim 20; Fig 100; 399pp; English.  
 XX The invention relates to an isolated secreted and transmembrane PRO

CC polypeptide having 80 % sequence identity to a sequence appearing  
 CC as ABG95851-ABG95934 or their associated signal peptide, or a sequence of  
 CC an extracellular domain of the proteins with their associated signal  
 CC peptide or lacking its associated signal peptide. Also included are  
 CC the nucleic acids encoding the proteins, vectors, host cells,  
 CC fusion proteins and antibodies which specifically bind to the proteins.  
 CC The proteins are useful for detecting a polypeptide designated as A, B, C  
 CC or D in a sample suspected of containing an A, B, C or D polypeptide,  
 CC by contacting the sample with a polypeptide designated as E, F, G, H or  
 CC I (or vice versa) and determining the formation of a A/E, B/F, B/G, C/H  
 CC or D/I polypeptide conjugate in the sample, where the formation of the  
 CC conjugate is indicative of the presence of an A, B, C or D polypeptide  
 CC in the sample, where A is a PRO10272 polypeptide, B is a PRO20110  
 CC polypeptide, C is a PRO10096 polypeptide, D is a PRO19760 polypeptide,  
 CC E is a PRO5801 polypeptide, F is a PRO1 polypeptide, G is a PRO20040  
 CC polypeptide, H is a PRO20233 polypeptide and I is a PRO1890  
 CC polypeptide. The sample comprises a cell suspected of expressing the A,  
 CC B, C or D polypeptide. The E, F, G, H or I polypeptide is labeled with  
 CC a detectable label or is attached to a solid support. The proteins are  
 CC useful for linking a bioactive molecule to a cell expressing a  
 CC polypeptide designated as A, B, C or D or E, F, G, H or I. The bioactive  
 CC molecule is a toxin, a radiolabel or an antibody. The bioactive molecule  
 CC causes death of the cell. A, B, C, D, E, F, G, H, or I, or antibodies  
 CC against them are useful for modulating a biological activity of a cell  
 CC expressing a polypeptide designated as A, B, C or D or E, F, G, H, or  
 CC I. The cell is killed. The proteins are useful for identifying  
 CC agonists or antagonists, for the preparation of a medicament useful in  
 CC the treatment of a condition which is responsive to the proteins, as  
 CC molecular weight markers for protein electrophoresis purposes, and as  
 CC therapeutic agents for treating sports-related joint problems,  
 CC articular cartilage defects, osteoarthritis or rheumatoid arthritis.  
 CC Nucleic acids encoding the proteins are useful as hybridisation probes,  
 CC in chromosome and gene mapping, in the generation of anti-sense RNA and  
 CC DNA, for the preparation of the proteins, to generate transgenic or  
 CC knockout animals which are useful in the development and screening of  
 CC therapeutic useful reagents, for chromosome identification, and in gene  
 CC therapy. The antibody is useful as a therapeutic agent, and in gene  
 CC assay and for affinity purification of the protein from recombinant  
 CC cell culture natural sources. The present sequence represents a novel  
 CC secreted or transmembrane protein of the invention.

XX Sequence 596 AA;

Query Match 2.0%; Score 8; DB 23; Length 596;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
 |||||  
 DB 39 SSGASTAT 46

## RESULT 49

ABU71266  
 ID ABU71266 standard; Protein; 596 AA.

XX AC ABU71266;

DT 10-JUN-2003 (first entry)

XX DE Human PRO1342 protein.

XX Human; PRO; secreted; transmembrane; cytostatic; TNF-alpha; blood;  
 KW tumour necrosis factor alpha release; chondrocyte cell; proliferation;  
 KW differentiation; tumour; gene therapy.

XX OS Homo sapiens.

XX US2003036143-A1.

XX 20-FEB-2003.

XX 02-JUL-2002; 2002US-0187600.

XX PR 16-SEP-1998; 98WO-US19330.  
PR 07-OCT-1998; 98WO-US21141.  
PR 01-DEC-1998; 98WO-US25108.  
PR 08-MAR-1999; 99WO-US05028.  
PR 14-MAY-1999; 99WO-US10733.  
PR 02-JUN-1999; 99WO-US12252.  
PR 01-SEP-1999; 99WO-US20111.  
PR 15-SEP-1999; 99WO-US21090.  
PR 01-DEC-1999; 99WO-US28301.  
PR 02-DEC-1999; 99WO-US28551.  
PR 30-DEC-1999; 99WO-US31274.  
PR 05-JAN-2000; 2000WO-US00219.  
PR 18-FEB-2000; 2000WO-US04341.  
PR 18-FEB-2000; 2000WO-US04342.  
PR 22-FEB-2000; 2000WO-US04414.  
PR 24-FEB-2000; 2000WO-US05004.  
PR 01-MAR-2000; 2000WO-US05601.  
PR 02-MAR-2000; 2000WO-US05841.  
PR 15-MAR-2000; 2000WO-US06884.  
PR 30-MAR-2000; 2000WO-US08439.  
PR 17-MAY-2000; 2000WO-US13705.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 30-MAY-2000; 2000WO-US14941.  
PR 02-JUN-2000; 2000WO-US15264.  
PR 28-JUL-2000; 2000WO-US20710.  
PR 24-AUG-2000; 2000WO-US23328.  
PR 08-NOV-2000; 2000WO-US30952.  
PR 01-DEC-2000; 2000WO-US32678.  
PR 20-DEC-2000; 2000WO-US34956.  
PR 28-FEB-2001; 2001WO-US05520.  
PR 01-JUN-2001; 2001WO-US17800.  
PR 20-JUN-2001; 2001WO-US19692.  
PR 29-JUN-2001; 2001WO-US21066.  
PR 09-JUL-2001; 2001WO-US21735.  
PR 29-AUG-2001; 2001WO-US27099.  
PR 18-SEP-1997; 97US-059263P.  
PR 18-SEP-1997; 97US-059266P.  
PR 17-OCT-1997; 97US-062250P.  
PR 21-OCT-1997; 97US-063486P.  
PR 24-OCT-1997; 97US-063120P.  
PR 24-OCT-1997; 97US-063121P.  
PR 28-OCT-1997; 97US-063540P.  
PR 28-OCT-1997; 97US-063541P.  
PR 28-OCT-1997; 97US-063544P.  
PR 28-OCT-1997; 97US-063564P.  
PR 29-OCT-1997; 97US-063734P.  
PR 31-OCT-1997; 97US-063870P.  
PR 31-OCT-1997; 97US-064103P.  
PR 13-NOV-1997; 97US-065311P.  
PR 21-NOV-1997; 97US-066120P.  
PR 24-NOV-1997; 97US-066466P.  
PR 24-NOV-1997; 97US-066772P.  
PR 11-DEC-1997; 97US-069335P.  
PR 12-DEC-1997; 97US-069425P.  
PR 17-DEC-1997; 97US-069870P.  
PR 18-DEC-1997; 97US-068017P.  
PR 10-MAR-1998; 98US-077450P.  
PR 11-MAR-1998; 98US-077632P.  
PR 11-MAR-1998; 98US-077649P.  
PR 20-MAR-1998; 98US-078886P.  
PR 20-MAR-1998; 98US-078939P.  
PR 27-MAR-1998; 98US-079664P.  
PR 27-MAR-1998; 98US-079786P.  
PR 31-MAR-1998; 98US-080107P.  
PR 31-MAR-1998; 98US-080194P.  
PR 01-APR-1998; 98US-080327P.  
PR 01-APR-1998; 98US-080333P.  
PR 08-APR-1998; 98US-081049P.  
PR 08-APR-1998; 98US-081070P.  
PR 09-APR-1998; 98US-081195P.  
PR 15-APR-1998; 98US-081838P.  
PR 21-APR-1998; 98US-082568P.  
PR 21-APR-1998; 98US-082569P.  
PR 22-APR-1998; 98US-082704P.  
PR 22-APR-1998; 98US-082797P.  
PR 28-APR-1998; 98US-083322P.  
PR 29-APR-1998; 98US-083495P.  
PR 29-APR-1998; 98US-083496P.  
PR 29-APR-1998; 98US-083499P.  
PR 29-APR-1998; 98US-083559P.  
PR 05-MAY-1998; 98US-084366P.  
PR 06-MAY-1998; 98US-084414P.  
PR 07-MAY-1998; 98US-084639P.  
PR 07-MAY-1998; 98US-084640P.  
PR 07-MAY-1998; 98US-084643P.  
PR 15-MAY-1998; 98US-085579P.  
PR 15-MAY-1998; 98US-085580P.  
PR 15-MAY-1998; 98US-085582P.  
PR 15-MAY-1998; 98US-085700P.  
PR 18-MAY-1998; 98US-086023P.  
PR 22-MAY-1998; 98US-086392P.  
PR 22-MAY-1998; 98US-086486P.  
PR 28-MAY-1998; 98US-087098P.  
PR 28-MAY-1998; 98US-087208P.  
PR 02-JUN-1998; 98US-087609P.  
PR 02-JUN-1998; 98US-087759P.  
PR 03-JUN-1998; 98US-087827P.  
PR 04-JUN-1998; 98US-088025P.  
PR 04-JUN-1998; 98US-088028P.  
PR 04-JUN-1998; 98US-088029P.  
PR 04-JUN-1998; 98US-088033P.  
PR 04-JUN-1998; 98US-088326P.  
PR 05-JUN-1998; 98US-088167P.  
PR 05-JUN-1998; 98US-088202P.  
PR 05-JUN-1998; 98US-088212P.  
PR 05-JUN-1998; 98US-088217P.  
PR 09-JUN-1998; 98US-088655P.  
PR 10-JUN-1998; 98US-088722P.  
PR 10-JUN-1998; 98US-088738P.  
PR 10-JUN-1998; 98US-088740P.  
PR 10-JUN-1998; 98US-088811P.  
PR 10-JUN-1998; 98US-088824P.  
PR 10-JUN-1998; 98US-088825P.  
PR 10-JUN-1998; 98US-088826P.  
PR 11-JUN-1998; 98US-088861P.  
PR 11-JUN-1998; 98US-088863P.  
PR 11-JUN-1998; 98US-088876P.  
PR 12-JUN-1998; 98US-089090P.  
PR 12-JUN-1998; 98US-089105P.  
PR 16-JUN-1998; 98US-089512P.  
PR 16-JUN-1998; 98US-089514P.  
PR 17-JUN-1998; 98US-089538P.  
PR 17-JUN-1998; 98US-089598P.  
PR 17-JUN-1998; 98US-089653P.  
PR 18-JUN-1998; 98US-089908P.  
PR 19-JUN-1998; 98US-089952P.  
PR 22-JUN-1998; 98US-090246P.  
PR 22-JUN-1998; 98US-090252P.  
PR 22-JUN-1998; 98US-090254P.  
PR 24-JUN-1998; 98US-090423P.  
PR 24-JUN-1998; 98US-090435P.  
PR 24-JUN-1998; 98US-090444P.  
PR 24-JUN-1998; 98US-090461P.  
PR 24-JUN-1998; 98US-090535P.  
PR 24-JUN-1998; 98US-090540P.  
PR 25-JUN-1998; 98US-090678P.  
PR 25-JUN-1998; 98US-090678P.  
PR 25-JUN-1998; 98US-090688P.  
PR 25-JUN-1998; 98US-090690P.  
PR 25-JUN-1998; 98US-090694P.  
PR 25-JUN-1998; 98US-090695P.  
PR 25-JUN-1998; 98US-090696P.  
PR 26-JUN-1998; 98US-090862P.  
PR 26-JUN-1998; 98US-090863P.  
PR 26-JUN-1998; 98US-091010P.





```

RESULT 51
ABU72001
ID ABU72001 standard; Protein; 596 AA.
XX
AC ABU72001;
XX
DT 11-JUN-2003 (first entry)
XX
DE Novel human secreted and transmembrane protein PRO1342.
XX
KW Human; secreted and transmembrane polypeptide;
KW chromosome mapping; gene mapping; transgenic animal; knockout animal;
KW therapeutic agent screening; chromosome identification; tissue typing;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN US2003018183-A1.
XX
PD 23-JAN-2003.
XX
PF 01-MAY-2002; 2002US-0063512.
XX
PR 06-DEC-2001; 2001US-0006867.
XX
PA (GETH ) GENENTECH INC.
XX
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;
XX
DR WPI; 2003-330984/31.
DR N-PSDB; ACA60412.
XX
XX
PT New secreted and transmembrane PRO polypeptides and nucleic acid
PT molecules encoding the polypeptides, useful in gene therapy or
PT preparing a medicament for treating a condition that is responsive to
PT the PRO polypeptide or antibody -
XX
PS Disclosure; Fig 100; 409pp; English.
XX
CC The invention describes novel isolated PRO polypeptides. The PRO
CC polypeptides or anti-PRO antibodies are useful in preparing a medicament
CC for treating a condition that is responsive to the PRO polypeptide or
CC antibody. The PRO nucleotide sequences may be used as hybridisation
CC probes in chromosome and gene mapping, or in generating antisense RNA
CC and DNA. PRO nucleic acids are also useful in preparing PRO polypeptides,
CC in assays to identify other proteins or molecules involved in binding
CC reaction, to generate transgenic animals or knockout animals, which in
CC turn are useful in the development and screening of therapeutically
CC useful reagents, for chromosome identification, and tissue typing. The
CC PRO polypeptides and nucleic acid molecules are also useful in gene
CC therapy, and as molecular weight markers for protein electrophoresis
CC purposes. The anti-PRO antibodies may be used in diagnostic assays for
CC PRO, or for the affinity purification of PRO from recombinant cell
CC culture or natural sources. This is the amino acid sequence of a novel
CC human secreted and transmembrane PRO polypeptide.
XX
SQ Sequence 596 AA;
Query Match 2.0%; Score 8; DB 24; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88
Db 39 SSGASTAT 46

RESULT 52
ABU72158
ID ABU72158 standard; Protein; 596 AA.
XX
AC ABU72158;
XX
DT 13-JUN-2003 (first entry)
XX
DE Human PRO polypeptide #50.
XX
KW Human; PRO polypeptide; secreted and transmembrane protein;
KW anti-PRO antibody; diagnostic assay; gene expression.
XX
OS Homo sapiens.
XX
PN US2003023042-A1.
XX
PD 30-JAN-2003.
XX
PF 01-MAY-2002; 2002US-0063502.
XX
PR 06-DEC-2001; 2001US-0006867.
XX
PA (GETH ) GENENTECH INC.
XX
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;
XX
DR WPI; 2003-331484/31.
DR N-PSDB; ACA63422.
XX
XX
PT Novel monoclonal antibody that binds to secreted and transmembrane
PT polypeptide, useful for detecting and purifying the polypeptide and
PT also for treating conditions responsive to the antibody -
XX
PS Disclosure; Fig 100; 408pp; English.
XX
CC The present invention relates to the isolation of novel human PRO
CC polypeptides and the polynucleotide sequences encoding them. The
CC PRO polypeptides are secreted and transmembrane proteins. The PRO
CC polypeptides and polynucleotides are useful for preparing a
CC medicament useful in the treatment of a condition responsive to
CC anti-PRO antibody. Anti-PRO antibodies are useful in diagnostic
CC assays for PRO, by detecting its expression in specific cells,
CC tissues or serum, and for affinity purification of PRO from
CC recombinant cell culture or natural sources. ABU72109-ABU72192
CC represent the human PRO polypeptides of the invention.
XX
SQ Sequence 596 AA;
Query Match 2.0%; Score 8; DB 24; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88
Db 39 SSGASTAT 46

RESULT 53
ABU65723
ID ABU65723 standard; Protein; 596 AA.
XX
AC ABU65723;
XX
DT 19-MAY-2003 (first entry)
XX
DE Human secreted/transmembrane protein, SEQ ID 310.
XX
KW Human; PRO; secreted protein; transmembrane protein;
KW cytostatic; antiarthritic; osteopathic; adrenal tumour; lung tumour;
KW colon tumour; breast tumour; prostate tumour; rectal tumour;
KW cervical tumour; liver tumour; TNF-alpha release; arthritis;
KW tumour necrosis factor alpha; chondrocyte cell; bone disorder;
KW cartilage disorder; sports injury.
XX
OS Homo sapiens.

```

XX US2003036156-A1.  
PN 02-JUL-2002; 2002US-0188767.  
XX 16-SEP-1998; 98WO-US19330.  
PR 07-OCT-1998; 98WO-US21141.  
PR 01-DEC-1998; 98WO-US25108.  
PR 08-MAR-1999; 99WO-US05028.  
PR 14-MAY-1999; 99WO-US10733.  
PR 02-JUN-1999; 99WO-US12252.  
PR 01-SEP-1999; 99WO-US20111.  
PR 15-SEP-1999; 99WO-US21090.  
PR 01-DEC-1999; 99WO-US28301.  
PR 02-DEC-1999; 99WO-US28551.  
PR 30-DEC-1999; 99WO-US31274.  
PR 05-JAN-2000; 2000WO-US00219.  
PR 18-FEB-2000; 2000WO-US04341.  
PR 18-FEB-2000; 2000WO-US04342.  
PR 22-FEB-2000; 2000WO-US04414.  
PR 24-FEB-2000; 2000WO-US05004.  
PR 01-MAR-2000; 2000WO-US05601.  
PR 02-MAR-2000; 2000WO-US05841.  
PR 15-MAR-2000; 2000WO-US06884.  
PR 30-MAR-2000; 2000WO-US08439.  
PR 17-MAY-2000; 2000WO-US13705.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 30-MAY-2000; 2000WO-US14941.  
PR 02-JUN-2000; 2000WO-US15264.  
PR 28-JUL-2000; 2000WO-US20710.  
PR 24-AUG-2000; 2000WO-US23328.  
PR 08-NOV-2000; 2000WO-US30952.  
PR 01-DEC-2000; 2000WO-US32678.  
PR 20-DEC-2000; 2000WO-US34956.  
PR 28-FEB-2001; 2001WO-US06520.  
PR 01-JUN-2001; 2001WO-US17800.  
PR 20-JUN-2001; 2001WO-US19692.  
PR 29-JUN-2001; 2001WO-US21066.  
PR 09-JUL-2001; 2001WO-US21735.  
PR 29-AUG-2001; 2001WO-US27099.  
PR 18-SEP-1997; 97US-US92631P.  
PR 18-SEP-1997; 97US-US92666P.  
PR 17-OCT-1997; 97US-US92250P.  
PR 21-OCT-1997; 97US-US93486P.  
PR 24-OCT-1997; 97US-US93120P.  
PR 24-OCT-1997; 97US-US93121P.  
PR 28-OCT-1997; 97US-US93540P.  
PR 28-OCT-1997; 97US-US93541P.  
PR 28-OCT-1997; 97US-US93544P.  
PR 28-OCT-1997; 97US-US93544P.  
PR 29-OCT-1997; 97US-US93734P.  
PR 31-OCT-1997; 97US-US93870P.  
PR 31-OCT-1997; 97US-US94103P.  
PR 13-NOV-1997; 97US-US95311P.  
PR 21-NOV-1997; 97US-US96130P.  
PR 24-NOV-1997; 97US-US96466P.  
PR 24-NOV-1997; 97US-US96772P.  
PR 11-DEC-1997; 97US-US99335P.  
PR 12-DEC-1997; 97US-US99425P.  
PR 17-DEC-1997; 97US-US99870P.  
PR 18-DEC-1997; 97US-US98017P.  
PR 10-MAR-1998; 98US-US97450P.  
PR 11-MAR-1998; 98US-US97632P.  
PR 11-MAR-1998; 98US-US97649P.  
PR 20-MAR-1998; 98US-US978886P.  
PR 20-MAR-1998; 98US-US978939P.  
PR 27-MAR-1998; 98US-US979664P.  
PR 27-MAR-1998; 98US-US979786P.  
PR 31-MAR-1998; 98US-US980107P.  
PR 31-MAR-1998; 98US-US980194P.  
PR 01-APR-1998; 98US-US980327P.  
PR 01-APR-1998; 98US-US980333P.  
PR 08-APR-1998; 98US-US98049P.  
PR 08-APR-1998; 98US-US981070P.  
PR 09-APR-1998; 98US-US981195P.  
PR 15-APR-1998; 98US-US981838P.  
PR 21-APR-1998; 98US-US982568P.  
PR 21-APR-1998; 98US-US982569P.  
PR 22-APR-1998; 98US-US982704P.  
PR 22-APR-1998; 98US-US982797P.  
PR 28-APR-1998; 98US-US983322P.  
PR 29-APR-1998; 98US-US983495P.  
PR 29-APR-1998; 98US-US983496P.  
PR 29-APR-1998; 98US-US983499P.  
PR 29-APR-1998; 98US-US983559P.  
PR 05-MAY-1998; 98US-US984366P.  
PR 06-MAY-1998; 98US-US984414P.  
PR 07-MAY-1998; 98US-US984639P.  
PR 07-MAY-1998; 98US-US984640P.  
PR 07-MAY-1998; 98US-US984643P.  
PR 15-MAY-1998; 98US-US985579P.  
PR 15-MAY-1998; 98US-US985580P.  
PR 15-MAY-1998; 98US-US985582P.  
PR 15-MAY-1998; 98US-US985700P.  
PR 18-MAY-1998; 98US-US986023P.  
PR 22-MAY-1998; 98US-US986392P.  
PR 22-MAY-1998; 98US-US986486P.  
PR 28-MAY-1998; 98US-US987098P.  
PR 28-MAY-1998; 98US-US987208P.  
PR 02-JUN-1998; 98US-US987609P.  
PR 02-JUN-1998; 98US-US987759P.  
PR 03-JUN-1998; 98US-US987827P.  
PR 04-JUN-1998; 98US-US988025P.  
PR 04-JUN-1998; 98US-US988028P.  
PR 04-JUN-1998; 98US-US988029P.  
PR 04-JUN-1998; 98US-US988033P.  
PR 04-JUN-1998; 98US-US988326P.  
PR 05-JUN-1998; 98US-US988167P.  
PR 05-JUN-1998; 98US-US988202P.  
PR 05-JUN-1998; 98US-US988212P.  
PR 05-JUN-1998; 98US-US988217P.  
PR 09-JUN-1998; 98US-US988555P.  
PR 10-JUN-1998; 98US-US988722P.  
PR 10-JUN-1998; 98US-US988738P.  
PR 10-JUN-1998; 98US-US988740P.  
PR 10-JUN-1998; 98US-US988811P.  
PR 10-JUN-1998; 98US-US988824P.  
PR 10-JUN-1998; 98US-US988825P.  
PR 10-JUN-1998; 98US-US988826P.  
PR 11-JUN-1998; 98US-US988861P.  
PR 11-JUN-1998; 98US-US988863P.  
PR 11-JUN-1998; 98US-US988876P.  
PR 12-JUN-1998; 98US-US989090P.  
PR 12-JUN-1998; 98US-US989105P.  
PR 16-JUN-1998; 98US-US989512P.  
PR 16-JUN-1998; 98US-US989514P.  
PR 17-JUN-1998; 98US-US989538P.  
PR 17-JUN-1998; 98US-US989598P.  
PR 17-JUN-1998; 98US-US989598P.  
PR 18-JUN-1998; 98US-US989598P.  
PR 19-JUN-1998; 98US-US989598P.  
PR 22-JUN-1998; 98US-US989598P.  
PR 22-JUN-1998; 98US-US989598P.  
PR 22-JUN-1998; 98US-US989598P.  
PR 24-JUN-1998; 98US-US989598P.  
PR 24-JUN-1998; 98US-US989598P.  
PR 24-JUN-1998; 98US-US989598P.  
PR 24-JUN-1998; 98US-US989598P.  
PR 25-JUN-1998; 98US-US989598P.  
PR 25-JUN-1998; 98US-US989598P.  
PR 25-JUN-1998; 98US-US989598P.  
PR 25-JUN-1998; 98US-US989598P.

PR 25-JUN-1998; 98US-090594P.  
 PR 25-JUN-1998; 98US-090595P.  
 PR 25-JUN-1998; 98US-090596P.  
 PR 26-JUN-1998; 98US-090862P.  
 PR 26-JUN-1998; 98US-090863P.  
 PR 26-JUN-1998; 98US-091010P.  
 PR 01-JUL-1998; 98US-091159P.  
 PR 01-JUL-1998; 98US-0911544P.  
 PR 02-JUL-1998; 98US-0911478P.  
 PR 02-JUL-1998; 98US-091486P.  
 PR 02-JUL-1998; 98US-091626P.  
 PR 02-JUL-1998; 98US-091628P.  
 PR 02-JUL-1998; 98US-091632P.  
 PR 24-JUL-1998; 98US-094006P.  
 PR 04-AUG-1998; 98US-095282P.  
 PR 10-AUG-1998; 98US-095398P.  
 PR 10-AUG-1998; 98US-096012P.  
 PR 17-AUG-1998; 98US-096757P.  
 PR 17-AUG-1998; 98US-096766P.  
 PR 17-AUG-1998; 98US-096867P.  
 PR 17-AUG-1998; 98US-096891P.  
 PR 17-AUG-1998; 98US-096897P.  
 PR 18-AUG-1998; 98US-096949P.  
 PR 18-AUG-1998; 98US-096959P.  
 PR 18-AUG-1998; 98US-097022P.  
 PR 26-AUG-1998; 98US-097952P.  
 PR 26-AUG-1998; 98US-097954P.  
 PR 26-AUG-1998; 98US-097955P.  
 PR 26-AUG-1998; 98US-097971P.  
 PR 26-AUG-1998; 98US-097974P.  
 PR 26-AUG-1998; 98US-098014P.  
 PR 01-SEP-1998; 98US-098716P.  
 PR 01-SEP-1998; 98US-098723P.  
 PR 02-SEP-1998; 98US-098803P.  
 PR 02-SEP-1998; 98US-098821P.  
 PR 02-SEP-1998; 98US-098843P.  
 PR 09-SEP-1998; 98US-099602P.  
 PR 10-SEP-1998; 98US-099741P.

Query Match 2.0%; Score 8; DB 24; Length 596;

Best Local Similarity 100.0%; Pred. No. 1.6e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
 Db 39 SSGASTAT 46

RESULT 54

ID ABU66056  
 ID ABU66056 standard; Protein; 596 AA.

XX AC ABU66056;

XX DT 20-MAY-2003 (first entry)

XX DE Novel human secreted and transmembrane protein PRO1342.

XX KW Human; secreted protein; transmembrane protein; cytostatic;  
 gene therapy; TNF-Agostat-Alpha; chondrocyte stimulator; tumour;  
 adrenal tumour; lung tumour; colon tumour; breast tumour;  
 prostate tumour; rectal tumour; cervical tumour; liver tumour.

XX OS Homo sapiens.

XX PN US2003036157-A1.

XX PD 20-FEB-2003.

XX PF 02-JUL-2002; 2002US-0188769.

XX XX 16-SEP-1998; 98WO-US19330.

PR 07-OCT-1998; 98WO-US21141.

PR 01-DEC-1998; 98WO-US25108.  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 14-MAY-1999; 99WO-US10733.  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 15-SEP-1999; 99WO-US21090.  
 PR 01-DEC-1999; 99WO-US28301.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 30-DEC-1999; 99WO-US31274.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 22-FEB-2000; 2000WO-US04414.  
 PR 24-FEB-2000; 2000WO-US05004.  
 PR 01-MAR-2000; 2000WO-US05601.  
 PR 02-MAR-2000; 2000WO-US05841.  
 PR 15-MAR-2000; 2000WO-US06884.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 17-MAY-2000; 2000WO-US13705.  
 PR 22-MAY-2000; 2000WO-US14042.  
 PR 30-MAY-2000; 2000WO-US14941.  
 PR 02-JUN-2000; 2000WO-US15264.  
 PR 28-JUL-2000; 2000WO-US20710.  
 PR 24-AUG-2000; 2000WO-US23328.  
 PR 08-NOV-2000; 2000WO-US30952.  
 PR 01-DEC-2000; 2000WO-US32678.  
 PR 20-DEC-2000; 2000WO-US34956.  
 PR 28-FEB-2001; 2001WO-US06520.  
 PR 01-JUN-2001; 2001WO-US17800.  
 PR 20-JUN-2001; 2001WO-US19692.  
 PR 29-JUN-2001; 2001WO-US21066.  
 PR 09-JUL-2001; 2001WO-US21735.  
 PR 29-AUG-2001; 2001WO-US27099.  
 PR 18-SEP-1997; 97US-059263P.  
 PR 18-SEP-1997; 97US-059266P.  
 PR 17-OCT-1997; 97US-062250P.  
 PR 21-OCT-1997; 97US-063486P.  
 PR 24-OCT-1997; 97US-063120P.  
 PR 24-OCT-1997; 97US-063121P.  
 PR 28-OCT-1997; 97US-063540P.  
 PR 28-OCT-1997; 97US-063541P.  
 PR 28-OCT-1997; 97US-063564P.  
 PR 29-OCT-1997; 97US-063734P.  
 PR 31-OCT-1997; 97US-063870P.  
 PR 31-OCT-1997; 97US-064103P.  
 PR 13-NOV-1997; 97US-065311P.  
 PR 21-NOV-1997; 97US-066120P.  
 PR 24-NOV-1997; 97US-066466P.  
 PR 11-DEC-1997; 97US-066772P.  
 PR 12-DEC-1997; 97US-069335P.  
 PR 17-DEC-1997; 97US-069425P.  
 PR 18-DEC-1997; 97US-069870P.  
 PR 10-MAR-1998; 98US-068017P.  
 PR 11-MAR-1998; 98US-077632P.  
 PR 11-MAR-1998; 98US-077649P.  
 PR 20-MAR-1998; 98US-078886P.  
 PR 27-MAR-1998; 98US-078939P.  
 PR 27-MAR-1998; 98US-079664P.  
 PR 31-MAR-1998; 98US-079786P.  
 PR 31-MAR-1998; 98US-080107P.  
 PR 31-MAR-1998; 98US-080194P.  
 PR 01-APR-1998; 98US-080327P.  
 PR 01-APR-1998; 98US-080333P.  
 PR 08-APR-1998; 98US-081049P.  
 PR 08-APR-1998; 98US-081070P.  
 PR 09-APR-1998; 98US-081195P.  
 PR 15-APR-1998; 98US-081838P.  
 PR 21-APR-1998; 98US-082568P.  
 PR 21-APR-1998; 98US-082569P.  
 PR 22-APR-1998; 98US-082704P.  
 PR 22-APR-1998; 98US-082797P.

PR 28-APR-1998; 98US-083322P.  
PR 29-APR-1998; 98US-083495P.  
PR 29-APR-1998; 98US-083496P.  
PR 29-APR-1998; 98US-083499P.  
PR 29-APR-1998; 98US-083599P.  
PR 05-MAY-1998; 98US-084366P.  
PR 06-MAY-1998; 98US-084414P.  
PR 07-MAY-1998; 98US-084639P.  
PR 07-MAY-1998; 98US-084640P.  
PR 07-MAY-1998; 98US-084643P.  
PR 15-MAY-1998; 98US-085579P.  
PR 15-MAY-1998; 98US-085580P.  
PR 15-MAY-1998; 98US-085582P.  
PR 15-MAY-1998; 98US-085700P.  
PR 18-MAY-1998; 98US-086023P.  
PR 22-MAY-1998; 98US-086392P.  
PR 22-MAY-1998; 98US-086486P.  
PR 28-MAY-1998; 98US-087098P.  
PR 28-MAY-1998; 98US-087208P.  
PR 02-JUN-1998; 98US-087609P.  
PR 02-JUN-1998; 98US-087759P.  
PR 03-JUN-1998; 98US-087827P.  
PR 04-JUN-1998; 98US-088025P.  
PR 04-JUN-1998; 98US-088028P.  
PR 04-JUN-1998; 98US-088029P.  
PR 04-JUN-1998; 98US-088033P.  
PR 04-JUN-1998; 98US-088326P.  
PR 05-JUN-1998; 98US-088167P.  
PR 05-JUN-1998; 98US-088202P.  
PR 05-JUN-1998; 98US-088212P.  
PR 05-JUN-1998; 98US-088217P.  
PR 09-JUN-1998; 98US-088655P.  
PR 10-JUN-1998; 98US-088722P.  
PR 10-JUN-1998; 98US-088738P.  
PR 10-JUN-1998; 98US-088740P.  
PR 10-JUN-1998; 98US-088811P.  
PR 10-JUN-1998; 98US-088824P.  
PR 10-JUN-1998; 98US-088825P.  
PR 10-JUN-1998; 98US-088826P.  
PR 11-JUN-1998; 98US-088861P.  
PR 11-JUN-1998; 98US-088863P.  
PR 11-JUN-1998; 98US-088876P.  
PR 12-JUN-1998; 98US-089090P.  
PR 12-JUN-1998; 98US-089105P.  
PR 16-JUN-1998; 98US-089512P.  
PR 16-JUN-1998; 98US-089514P.  
PR 17-JUN-1998; 98US-089538P.  
PR 17-JUN-1998; 98US-089598P.  
PR 17-JUN-1998; 98US-089653P.  
PR 18-JUN-1998; 98US-089908P.  
PR 19-JUN-1998; 98US-089952P.  
PR 22-JUN-1998; 98US-090246P.  
PR 22-JUN-1998; 98US-090252P.  
PR 22-JUN-1998; 98US-090254P.  
PR 24-JUN-1998; 98US-090429P.  
PR 24-JUN-1998; 98US-090435P.  
PR 24-JUN-1998; 98US-090444P.  
PR 24-JUN-1998; 98US-090461P.  
PR 24-JUN-1998; 98US-090535P.  
PR 24-JUN-1998; 98US-090540P.  
PR 25-JUN-1998; 98US-090676P.  
PR 25-JUN-1998; 98US-090678P.  
PR 25-JUN-1998; 98US-090688P.  
PR 25-JUN-1998; 98US-090690P.  
PR 25-JUN-1998; 98US-090694P.  
PR 25-JUN-1998; 98US-090695P.  
PR 25-JUN-1998; 98US-090696P.  
PR 26-JUN-1998; 98US-090862P.  
PR 26-JUN-1998; 98US-090863P.  
PR 26-JUN-1998; 98US-091010P.  
PR 01-JUL-1998; 98US-091359P.  
PR 01-JUL-1998; 98US-091544P.  
PR 02-JUL-1998; 98US-091478P.

PR 02-JUL-1998; 98US-091486P.  
PR 02-JUL-1998; 98US-091626P.  
PR 02-JUL-1998; 98US-091628P.  
PR 02-JUL-1998; 98US-091632P.  
PR 24-JUL-1998; 98US-094006P.  
PR 04-AUG-1998; 98US-095282P.  
PR 10-AUG-1998; 98US-095998P.  
PR 10-AUG-1998; 98US-096012P.  
PR 17-AUG-1998; 98US-096757P.  
PR 17-AUG-1998; 98US-096766P.  
PR 17-AUG-1998; 98US-096867P.  
PR 17-AUG-1998; 98US-096891P.  
PR 17-AUG-1998; 98US-096897P.  
PR 18-AUG-1998; 98US-096949P.  
PR 18-AUG-1998; 98US-096959P.  
PR 18-AUG-1998; 98US-097022P.  
PR 26-AUG-1998; 98US-097952P.  
PR 26-AUG-1998; 98US-097954P.  
PR 26-AUG-1998; 98US-097955P.  
PR 26-AUG-1998; 98US-097971P.  
PR 26-AUG-1998; 98US-097974P.  
PR 26-AUG-1998; 98US-098014P.  
PR 01-SEP-1998; 98US-098716P.  
PR 01-SEP-1998; 98US-098723P.  
PR 02-SEP-1998; 98US-098803P.  
PR 02-SEP-1998; 98US-098821P.  
PR 02-SEP-1998; 98US-098843P.  
PR 09-SEP-1998; 98US-099602P.  
PR 10-SEP-1998; 98US-099741P.  
PR 10-SEP-1998; 98US-099754P.  
PR 10-SEP-1998; 98US-099763P.

Query Match 2.0%; Score 8; DB 24; Length 596;

Best Local Similarity 100.0%; Pred.No. 1.6e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88

Db 39 SSGASTAT 46

RESULT 55

ABU67560

ID ABU67560 standard; Protein; 596 AA.

XX AC ABU67560;

XX AC ABU67560;

DT 29-MAY-2003 (first entry)

XX Human secreted/transmembrane protein (PRO) #155.

XX Human; secreted and transmembrane protein; PRO; TNF-alpha;  
tumour necrosis factor alpha; chondrocyte cell; tumour; gene therapy;  
tissue typing.

XX Homo sapiens.

OS US2003036162-A1.

PN 20-FEB-2003.

XX 12-JUL-2002; 2002US-0194423.

XX 16-SEP-1998; 98WO-US19330.

PR 07-OCT-1998; 98WO-US21141.

PR 01-DEC-1998; 98WO-US25108.

PR 08-MAR-1999; 99WO-US05028.

PR 14-MAY-1999; 99WO-US10733.

PR 02-JUN-1999; 99WO-US12252.

PR 01-SEP-1999; 99WO-US20111.

PR 15-SEP-1999; 99WO-US21090.

PR 01-DEC-1999; 99WO-US28301.

PR 02-DEC-1999; 99WO-US28551.

PR 30-DEC-1999; 99WO-US31274.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 22-FEB-2000; 2000WO-US04414.  
 PR 24-FEB-2000; 2000WO-US05004.  
 PR 01-MAR-2000; 2000WO-US05601.  
 PR 02-MAR-2000; 2000WO-US05841.  
 PR 15-MAR-2000; 2000WO-US06884.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 17-MAY-2000; 2000WO-US13705.  
 PR 22-MAY-2000; 2000WO-US14042.  
 PR 30-MAY-2000; 2000WO-US14941.  
 PR 02-JUN-2000; 2000WO-US15264.  
 PR 28-JUL-2000; 2000WO-US20710.  
 PR 24-AUG-2000; 2000WO-US23328.  
 PR 08-NOV-2000; 2000WO-US30952.  
 PR 01-DEC-2000; 2000WO-US32678.  
 PR 28-DEC-2000; 2000WO-US34956.  
 PR 28-FEB-2001; 2001WO-US06520.  
 PR 01-JUN-2001; 2001WO-US17800.  
 PR 20-JUN-2001; 2001WO-US19692.  
 PR 29-JUN-2001; 2001WO-US21066.  
 PR 09-JUL-2001; 2001WO-US21735.  
 PR 28-AUG-2001; 2001WO-US27099.  
 PR 26-JUN-1998; 98US-0105413.  
 PR 07-OCT-1998; 98US-0168978.  
 PR 06-NOV-1998; 98US-0187368.  
 PR 07-DEC-1998; 98US-0202054.  
 PR 03-MAR-1999; 98US-0254311.  
 PR 14-MAY-1999; 98US-0311832.  
 PR 14-MAY-1999; 98US-0380137.  
 PR 25-AUG-1999; 98US-0380138.  
 PR 25-AUG-1999; 98US-0380139.  
 PR 25-AUG-1999; 98US-0380142.  
 PR 18-OCT-1999; 98US-0403297.  
 PR 12-NOV-1999; 98US-0423844.  
 PR 22-AUG-2000; 2000US-0644848.  
 PR 18-SEP-2000; 2000US-0664610.  
 PR 18-SEP-2000; 2000US-0665350.  
 PR 08-NOV-2000; 2000US-0709238.  
 PR 20-DEC-2000; 2000US-0747259.  
 PR 22-MAR-2001; 2001US-0816744.  
 PR 10-MAY-2001; 2001US-0854208.  
 PR 10-MAY-2001; 2001US-0854280.  
 PR 05-JUN-2001; 2001US-0866028.  
 PR 05-JUN-2001; 2001US-0874503.  
 PR 18-JUL-2001; 2001US-0908827.  
 PR 30-JUL-2001; 2001US-0918585.  
 PR 06-AUG-2001; 2001US-0924419.  
 PR 13-AUG-2001; 2001US-0929404.  
 PR 16-AUG-2001; 2001US-0931836.  
 PR 28-AUG-2001; 2001US-0941992.  
 PR 04-SEP-2001; 2001US-0946374.  
 PR 15-JAN-2002; 2002US-0052586.

(GETH) GENENTECH INC.

Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
 Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
 WPI; 2003-332039/31.  
 N-PSDB; ACA05854.

New secreted and transmembrane PRO polypeptides and nucleic acids,  
 useful in gene therapy, in chromosome and gene mapping, as chromosome  
 markers, in tissue typing, and in chromosome identification -

Claim 11; Fig 310; 706pp; English.

The invention discloses human nucleic acids encoding secreted and  
 transmembrane (PRO) polypeptides. Also disclosed is an antibody that  
 specifically binds to the PRO polypeptide, a method for stimulating the

CC release of tumour necrosis factor alpha (TNF-alpha) from human blood by  
 CC contacting the blood a PRO polypeptide, a method for stimulating the  
 CC proliferation or differentiation of chondrocyte cells by contacting the  
 CC cells with a PRO polypeptide, a method for detecting the presence of a  
 CC tumour in a mammal and an oligonucleotide probe derived from any of the  
 CC PRO nucleotide sequences. The nucleotide sequences are useful as probes,  
 CC in chromosome and gene mapping, in generating antisense RNA and DNA, in  
 CC preparing PRO polypeptides by recombinant techniques and in gene therapy  
 CC (e.g. for replacement of defective gene). The PRO polypeptides are useful  
 CC as molecular weight markers for protein electrophoresis purposes, for  
 CC chromosome identification, as chromosome markers, as therapeutic agents,  
 CC for stimulating the release of TNF-alpha from human blood, for  
 CC stimulating the proliferation or differentiation of chondrocytes and  
 CC detecting the presence of a tumour. The PRO polypeptides and nucleic  
 CC acids may also be used diagnostically for tissue typing. The sequences  
 CC presented in ABU67406-ABU67710 are the PRO polypeptides of the invention.

XX Sequence 596 AA;

Query Match 2.0%; Score 8; DB 24; Length 596;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
 Db 39 SSGASTAT 46  
 |||||

RESULT 56

ID ABU65418 standard; Protein; 596 AA.

XX AC ABU65418;

XX DT 16-MAY-2003 (first entry)

XX DE Human PRO polypeptide #155.

XX KW Human; PRO; cytostatic; chromosome mapping; gene mapping;  
 KW protein electrophoresis; tumour necrosis factor-alpha; TNF-alpha; blood;  
 XX chondrocyte differentiation; chondrocyte proliferation; tumour.

XX OS Homo sapiens.

XX FN US2003032102-A1.

XX PD 13-FEB-2003.

XX PF 17-JUN-2002; 2002US-0173697.

XX PR 16-SEP-1998; 98WO-US19330.

PR 07-OCT-1998; 98WO-US21141.

PR 01-DEC-1998; 98WO-US25108.

PR 08-MAR-1999; 99WO-US05028.

PR 14-MAY-1999; 99WO-US10733.

PR 02-JUN-1999; 99WO-US12252.

PR 01-SEP-1999; 99WO-US20111.

PR 15-SEP-1999; 99WO-US21090.

PR 01-DEC-1999; 99WO-US28301.

PR 02-DEC-1999; 99WO-US28551.

PR 30-DEC-1999; 99WO-US31274.

PR 05-JAN-2000; 2000WO-US00219.

PR 18-FEB-2000; 2000WO-US04341.

PR 22-FEB-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.

PR 02-MAR-2000; 2000WO-US05841.

PR 15-MAR-2000; 2000WO-US06884.

PR 30-MAR-2000; 2000WO-US08439.

PR 17-MAY-2000; 2000WO-US13705.

PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000WO-US14941.

PR 02-JUN-2000; 2000WO-US15264.  
PR 28-JUL-2000; 2000WO-US20710.  
PR 24-AUG-2000; 2000WO-US23328.  
PR 08-NOV-2000; 2000WO-US30352.  
PR 01-DEC-2000; 2000WO-US32578.  
PR 20-DEC-2000; 2000WO-US34956.  
PR 28-FEB-2001; 2001WO-US06520.  
PR 01-JUN-2001; 2001WO-US17800.  
PR 20-JUN-2001; 2001WO-US19692.  
PR 29-JUN-2001; 2001WO-US21066.  
PR 08-JUL-2001; 2001WO-US21735.  
PR 29-AUG-2001; 2001WO-US27099.  
PR 18-SEP-1997; 97US-059263P.  
PR 18-SEP-1997; 97US-059266P.  
PR 17-OCT-1997; 97US-062250P.  
PR 21-OCT-1997; 97US-063486P.  
PR 24-OCT-1997; 97US-063120P.  
PR 24-OCT-1997; 97US-063121P.  
PR 28-OCT-1997; 97US-063540P.  
PR 28-OCT-1997; 97US-063541P.  
PR 28-OCT-1997; 97US-063544P.  
PR 28-OCT-1997; 97US-063564P.  
PR 29-OCT-1997; 97US-063734P.  
PR 31-OCT-1997; 97US-063870P.  
PR 31-OCT-1997; 97US-064103P.  
PR 13-NOV-1997; 97US-065311P.  
PR 21-NOV-1997; 97US-066120P.  
PR 24-NOV-1997; 97US-066466P.  
PR 24-NOV-1997; 97US-066772P.  
PR 11-DEC-1997; 97US-069335P.  
PR 12-DEC-1997; 97US-069425P.  
PR 17-DEC-1997; 97US-069870P.  
PR 18-DEC-1997; 97US-068017P.  
PR 10-MAR-1998; 98US-077450P.  
PR 11-MAR-1998; 98US-077632P.  
PR 11-MAR-1998; 98US-077649P.  
PR 20-MAR-1998; 98US-078886P.  
PR 27-MAR-1998; 98US-078939P.  
PR 27-MAR-1998; 98US-079664P.  
PR 31-MAR-1998; 98US-079786P.  
PR 31-MAR-1998; 98US-080107P.  
PR 31-MAR-1998; 98US-080194P.  
PR 01-APR-1998; 98US-080327P.  
PR 01-APR-1998; 98US-080333P.  
PR 08-APR-1998; 98US-081049P.  
PR 08-APR-1998; 98US-081070P.  
PR 15-APR-1998; 98US-081195P.  
PR 15-APR-1998; 98US-081838P.  
PR 21-APR-1998; 98US-082569P.  
PR 21-APR-1998; 98US-082569P.  
PR 22-APR-1998; 98US-082704P.  
PR 22-APR-1998; 98US-082797P.  
PR 28-APR-1998; 98US-083322P.  
PR 29-APR-1998; 98US-083495P.  
PR 29-APR-1998; 98US-083496P.  
PR 29-APR-1998; 98US-083499P.  
PR 29-APR-1998; 98US-083559P.  
PR 05-MAY-1998; 98US-084366P.  
PR 08-MAY-1998; 98US-084414P.  
PR 07-MAY-1998; 98US-084639P.  
PR 07-MAY-1998; 98US-084640P.  
PR 07-MAY-1998; 98US-084643P.  
PR 15-MAY-1998; 98US-085579P.  
PR 15-MAY-1998; 98US-085580P.  
PR 15-MAY-1998; 98US-085582P.  
PR 15-MAY-1998; 98US-085700P.  
PR 18-MAY-1998; 98US-086023P.  
PR 22-MAY-1998; 98US-086392P.  
PR 22-MAY-1998; 98US-086466P.  
PR 28-MAY-1998; 98US-087098P.  
PR 28-MAY-1998; 98US-087208P.  
PR 02-JUN-1998; 98US-087609P.  
PR 02-JUN-1998; 98US-087759P.  
PR 03-JUN-1998; 98US-087827P.  
PR 04-JUN-1998; 98US-088025P.  
PR 04-JUN-1998; 98US-088028P.  
PR 04-JUN-1998; 98US-088029P.  
PR 04-JUN-1998; 98US-088033P.  
PR 04-JUN-1998; 98US-088126P.  
PR 05-JUN-1998; 98US-088167P.  
PR 05-JUN-1998; 98US-088202P.  
PR 05-JUN-1998; 98US-088212P.  
PR 05-JUN-1998; 98US-088217P.  
PR 09-JUN-1998; 98US-088655P.  
PR 10-JUN-1998; 98US-088722P.  
PR 10-JUN-1998; 98US-088738P.  
PR 10-JUN-1998; 98US-088740P.  
PR 10-JUN-1998; 98US-088811P.  
PR 10-JUN-1998; 98US-088824P.  
PR 10-JUN-1998; 98US-088825P.  
PR 10-JUN-1998; 98US-088826P.  
PR 11-JUN-1998; 98US-088861P.  
PR 11-JUN-1998; 98US-088863P.  
PR 11-JUN-1998; 98US-088876P.  
PR 12-JUN-1998; 98US-089090P.  
PR 12-JUN-1998; 98US-089105P.  
PR 16-JUN-1998; 98US-089512P.  
PR 16-JUN-1998; 98US-089514P.  
PR 17-JUN-1998; 98US-089538P.  
PR 17-JUN-1998; 98US-089598P.  
PR 17-JUN-1998; 98US-089653P.  
PR 18-JUN-1998; 98US-089908P.  
PR 19-JUN-1998; 98US-089952P.  
PR 22-JUN-1998; 98US-090246P.  
PR 22-JUN-1998; 98US-090254P.  
PR 24-JUN-1998; 98US-090429P.  
PR 24-JUN-1998; 98US-090435P.  
PR 24-JUN-1998; 98US-090444P.  
PR 24-JUN-1998; 98US-090461P.  
PR 24-JUN-1998; 98US-090535P.  
PR 24-JUN-1998; 98US-090540P.  
PR 25-JUN-1998; 98US-090676P.  
PR 25-JUN-1998; 98US-090678P.  
PR 25-JUN-1998; 98US-090688P.  
PR 25-JUN-1998; 98US-090690P.  
PR 25-JUN-1998; 98US-090694P.  
PR 25-JUN-1998; 98US-090695P.  
PR 25-JUN-1998; 98US-090696P.  
PR 26-JUN-1998; 98US-090862P.  
PR 26-JUN-1998; 98US-090863P.  
PR 01-JUL-1998; 98US-091010P.  
PR 01-JUL-1998; 98US-091359P.  
PR 01-JUL-1998; 98US-091544P.  
PR 02-JUL-1998; 98US-091478P.  
PR 02-JUL-1998; 98US-091486P.  
PR 02-JUL-1998; 98US-091626P.  
PR 02-JUL-1998; 98US-091628P.  
PR 02-JUL-1998; 98US-091632P.  
PR 24-JUL-1998; 98US-094006P.  
PR 04-AUG-1998; 98US-095282P.  
PR 10-AUG-1998; 98US-095998P.  
PR 10-AUG-1998; 98US-096012P.  
PR 17-AUG-1998; 98US-096757P.  
PR 17-AUG-1998; 98US-096766P.  
PR 17-AUG-1998; 98US-096867P.  
PR 17-AUG-1998; 98US-096891P.  
PR 17-AUG-1998; 98US-096897P.  
PR 18-AUG-1998; 98US-096949P.  
PR 18-AUG-1998; 98US-096959P.  
PR 26-AUG-1998; 98US-097022P.  
PR 26-AUG-1998; 98US-097952P.  
PR 26-AUG-1998; 98US-097954P.  
PR 26-AUG-1998; 98US-097955P.  
PR 26-AUG-1998; 98US-097971P.  
PR 26-AUG-1998; 98US-097974P.

PR 26-AUG-1998; 98US-098014P.  
 PR 01-SEP-1998; 98US-098716P.  
 PR 01-SEP-1998; 98US-098723P.  
 PR 02-SEP-1998; 98US-098803P.  
 PR 02-SEP-1998; 98US-098821P.  
 PR 02-SEP-1998; 98US-098843P.  
 PR 09-SEP-1998; 98US-099602P.  
 PR 10-SEP-1998; 98US-099741P.  
 PR 10-SEP-1998; 98US-099754P.  
 PR 10-SEP-1998; 98US-099763P.  
 PR 10-SEP-1998; 98US-099812P.

Query Match 2.0%; Score 8; DB 24; Length 596;

Best Local Similarity 100.0%; Pred. NO. 1.6e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 81 SSGASTAT 88

Db 39 SSGASTAT 46

RESULT 57

ABUS8554

ID ABUS8554 standard; Protein; 596 AA.

XX AC ABUS8554;

XX DT 15-APR-2003 (first entry)

XX DE Human PRO polypeptide #155.

XX KW Human; PRO; cytostatic; tumour; cancer; lung; stomach;  
 KW liver; dog; cat; cow; horse; sheep; pig; goat; rabbit; ADEPT;  
 KW antibody-dependent enzyme mediated prodrug therapy.

XX OS Homo sapiens.

XX PN US2003027272-A1.

XX PD 06-FEB-2003.

XX PF 21-JUN-2002; 2002US-0176492.

XX 16-SEP-1998; 98WO-US19330.  
 PR 07-OCT-1998; 98WO-US21141.  
 PR 01-DEC-1998; 98WO-US25108.  
 PR 08-MAR-1999; 98WO-US05028.  
 PR 10-MAR-1999; 98WO-US05190.  
 PR 14-MAY-1999; 98WO-US10733.  
 PR 02-JUN-1999; 98WO-US12252.  
 PR 01-SEP-1999; 98WO-US20111.  
 PR 15-SEP-1999; 98WO-US21090.  
 PR 30-NOV-1999; 98WO-US28313.  
 PR 01-DEC-1999; 98WO-US28301.  
 PR 02-DEC-1999; 98WO-US28551.  
 PR 05-JAN-2000; 98WO-US31274.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 22-FEB-2000; 2000WO-US04414.  
 PR 24-FEB-2000; 2000WO-US05004.  
 PR 01-MAR-2000; 2000WO-US05061.  
 PR 02-MAR-2000; 2000WO-US05841.  
 PR 10-MAR-2000; 2000WO-US06319.  
 PR 15-MAR-2000; 2000WO-US06884.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 17-MAY-2000; 2000WO-US13705.  
 PR 22-MAY-2000; 2000WO-US14042.  
 PR 30-MAY-2000; 2000WO-US14941.  
 PR 02-JUN-2000; 2000WO-US15264.  
 PR 28-JUL-2000; 2000WO-US20710.  
 PR 24-AUG-2000; 2000WO-US23328.  
 PR 08-NOV-2000; 2000WO-US30952.

PR 10-NOV-2000; 2000WO-US30873.  
 PR 01-DEC-2000; 2000WO-US32678.  
 PR 20-DEC-2000; 2000WO-US34956.  
 PR 28-FEB-2001; 2001WO-US06520.  
 PR 01-JUN-2001; 2001WO-US17800.  
 PR 20-JUN-2001; 2001WO-US19692.  
 PR 29-JUN-2001; 2001WO-US21066.  
 PR 09-JUL-2001; 2001WO-US21735.  
 PR 18-SEP-1997; 97US-059263P.  
 PR 18-SEP-1997; 97US-059266P.  
 PR 17-OCT-1997; 97US-062250P.  
 PR 21-OCT-1997; 97US-063486P.  
 PR 24-OCT-1997; 97US-063120P.  
 PR 28-OCT-1997; 97US-063121P.  
 PR 28-OCT-1997; 97US-063540P.  
 PR 28-OCT-1997; 97US-063541P.  
 PR 28-OCT-1997; 97US-063544P.  
 PR 28-OCT-1997; 97US-063564P.  
 PR 29-OCT-1997; 97US-063734P.  
 PR 31-OCT-1997; 97US-063870P.  
 PR 31-OCT-1997; 97US-064103P.  
 PR 13-NOV-1997; 97US-065311P.  
 PR 21-NOV-1997; 97US-066120P.  
 PR 24-NOV-1997; 97US-066466P.  
 PR 11-DEC-1997; 97US-069335P.  
 PR 12-DEC-1997; 97US-069425P.  
 PR 17-DEC-1997; 97US-069870P.  
 PR 18-DEC-1997; 97US-068017P.  
 PR 10-MAR-1998; 98US-077450P.  
 PR 11-MAR-1998; 98US-077632P.  
 PR 11-MAR-1998; 98US-077649P.  
 PR 20-MAR-1998; 98US-078886P.  
 PR 20-MAR-1998; 98US-078939P.  
 PR 27-MAR-1998; 98US-079664P.  
 PR 27-MAR-1998; 98US-079786P.  
 PR 31-MAR-1998; 98US-080194P.  
 PR 01-APR-1998; 98US-080327P.  
 PR 01-APR-1998; 98US-080333P.  
 PR 08-APR-1998; 98US-081049P.  
 PR 08-APR-1998; 98US-081070P.  
 PR 09-APR-1998; 98US-081195P.  
 PR 15-APR-1998; 98US-081838P.  
 PR 21-APR-1998; 98US-082568P.  
 PR 21-APR-1998; 98US-082569P.  
 PR 22-APR-1998; 98US-082704P.  
 PR 22-APR-1998; 98US-082797P.  
 PR 29-APR-1998; 98US-083322P.  
 PR 29-APR-1998; 98US-083495P.  
 PR 29-APR-1998; 98US-083496P.  
 PR 29-APR-1998; 98US-083499P.  
 PR 29-APR-1998; 98US-083559P.  
 PR 05-MAY-1998; 98US-084366P.  
 PR 06-MAY-1998; 98US-084414P.  
 PR 07-MAY-1998; 98US-084639P.  
 PR 07-MAY-1998; 98US-084640P.  
 PR 07-MAY-1998; 98US-084643P.  
 PR 15-MAY-1998; 98US-085579P.  
 PR 15-MAY-1998; 98US-085580P.  
 PR 15-MAY-1998; 98US-085582P.  
 PR 15-MAY-1998; 98US-085700P.  
 PR 18-MAY-1998; 98US-086023P.  
 PR 22-MAY-1998; 98US-086392P.  
 PR 22-MAY-1998; 98US-086486P.  
 PR 28-MAY-1998; 98US-087098P.  
 PR 28-MAY-1998; 98US-087208P.  
 PR 02-JUN-1998; 98US-087609P.  
 PR 02-JUN-1998; 98US-087759P.  
 PR 03-JUN-1998; 98US-087827P.  
 PR 04-JUN-1998; 98US-088025P.  
 PR 04-JUN-1998; 98US-088028P.  
 PR 04-JUN-1998; 98US-088029P.





PR 05-JAN-2000; 2000WO-US00219.  
PR 06-JAN-2000; 2000WO-US00277.  
PR 06-JAN-2000; 2000WO-US00376.  
PR 11-FEB-2000; 2000WO-US03565.  
PR 18-FEB-2000; 2000WO-US04341.  
PR 18-FEB-2000; 2000WO-US04342.  
PR 22-FEB-2000; 2000WO-US04414.  
PR 24-FEB-2000; 2000WO-US05004.  
PR 01-MAR-2000; 2000WO-US05601.  
PR 02-MAR-2000; 2000WO-US05841.  
PR 15-MAR-2000; 2000WO-US06319.  
PR 21-MAR-2000; 2000WO-US06884.  
PR 30-MAR-2000; 2000WO-US07532.  
PR 17-MAY-2000; 2000WO-US08439.  
PR 22-MAY-2000; 2000WO-US13705.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 30-MAY-2000; 2000WO-US14941.  
PR 02-JUN-2000; 2000WO-US15264.  
PR 28-JUL-2000; 2000WO-US20710.  
PR 11-AUG-2000; 2000WO-US22031.  
PR 23-AUG-2000; 2000WO-US23322.  
PR 24-AUG-2000; 2000WO-US23328.  
PR 08-NOV-2000; 2000WO-US30952.  
PR 10-NOV-2000; 2000WO-US30873.  
PR 01-DEC-2000; 2000WO-US32678.  
PR 20-DEC-2000; 2000WO-US34956.  
PR 28-FEB-2001; 2001WO-US06520.  
PR 22-MAR-2001; 2001WO-US09552.  
PR 25-MAY-2001; 2001WO-US17092.  
PR 01-JUN-2001; 2001WO-US17800.  
PR 20-JUN-2001; 2001WO-US19692.  
PR 29-JUN-2001; 2001WO-US21066.  
PR 09-JUL-2001; 2001WO-US21735.  
PR 29-AUG-2001; 2001WO-US27099.  
PR 18-SEP-1997; 97US-US9263P.  
PR 18-SEP-1997; 97US-US9266P.  
PR 17-OCT-1997; 97US-US92250P.  
PR 21-OCT-1997; 97US-US92486P.  
PR 24-OCT-1997; 97US-US93120P.  
PR 24-OCT-1997; 97US-US93121P.  
PR 28-OCT-1997; 97US-US93540P.  
PR 28-OCT-1997; 97US-US93541P.  
PR 28-OCT-1997; 97US-US93544P.  
PR 28-OCT-1997; 97US-US93564P.  
PR 29-OCT-1997; 97US-US93734P.  
PR 31-OCT-1997; 97US-US93870P.  
PR 31-OCT-1997; 97US-US94103P.  
PR 13-NOV-1997; 97US-US95311P.  
PR 21-NOV-1997; 97US-US966120P.  
PR 24-NOV-1997; 97US-US966466P.  
PR 24-NOV-1997; 97US-US96772P.  
PR 11-DEC-1997; 97US-US99335P.  
PR 12-DEC-1997; 97US-US99425P.  
PR 17-DEC-1997; 97US-US99870P.  
PR 18-DEC-1997; 97US-US98017P.  
PR 10-MAR-1998; 98US-US97450P.  
PR 11-MAR-1998; 98US-US97632P.  
PR 20-MAR-1998; 98US-US97886P.  
PR 20-MAR-1998; 98US-US97893P.  
PR 27-MAR-1998; 98US-US979664P.  
PR 27-MAR-1998; 98US-US979786P.  
PR 31-MAR-1998; 98US-US980107P.  
PR 31-MAR-1998; 98US-US980194P.  
PR 01-APR-1998; 98US-US980327P.  
PR 01-APR-1998; 98US-US980333P.  
PR 08-APR-1998; 98US-US981049P.  
PR 08-APR-1998; 98US-US981070P.  
PR 09-APR-1998; 98US-US981195P.  
PR 15-APR-1998; 98US-US981838P.  
PR 21-APR-1998; 98US-US982568P.  
PR 21-APR-1998; 98US-US982569P.  
PR 22-APR-1998; 98US-US982704P.

PR 22-APR-1998; 98US-US982797P.  
PR 28-APR-1998; 98US-US983322P.  
PR 29-APR-1998; 98US-US983495P.  
PR 29-APR-1998; 98US-US983496P.  
PR 29-APR-1998; 98US-US983499P.  
PR 29-APR-1998; 98US-US983559P.  
PR 05-MAY-1998; 98US-US98366P.  
PR 06-MAY-1998; 98US-US984414P.  
PR 07-MAY-1998; 98US-US984640P.  
PR 07-MAY-1998; 98US-US984640P.  
PR 15-MAY-1998; 98US-US985579P.  
PR 15-MAY-1998; 98US-US985580P.  
PR 15-MAY-1998; 98US-US985582P.  
PR 15-MAY-1998; 98US-US985700P.  
PR 18-MAY-1998; 98US-US986023P.  
PR 22-MAY-1998; 98US-US986192P.  
PR 22-MAY-1998; 98US-US986486P.  
PR 28-MAY-1998; 98US-US987098P.  
PR 28-MAY-1998; 98US-US987208P.  
PR 02-JUN-1998; 98US-US987609P.  
PR 02-JUN-1998; 98US-US987759P.  
PR 03-JUN-1998; 98US-US987827P.  
PR 04-JUN-1998; 98US-US988025P.  
PR 04-JUN-1998; 98US-US988028P.  
PR 04-JUN-1998; 98US-US988029P.  
PR 04-JUN-1998; 98US-US988033P.  
PR 04-JUN-1998; 98US-US988326P.  
PR 05-JUN-1998; 98US-US988167P.  
PR 05-JUN-1998; 98US-US988202P.  
PR 05-JUN-1998; 98US-US988212P.  
PR 05-JUN-1998; 98US-US988217P.  
PR 09-JUN-1998; 98US-US988655P.  
PR 10-JUN-1998; 98US-US988722P.  
PR 10-JUN-1998; 98US-US98738P.  
PR 10-JUN-1998; 98US-US98740P.  
PR 10-JUN-1998; 98US-US98811P.  
PR 10-JUN-1998; 98US-US98824P.  
PR 10-JUN-1998; 98US-US98825P.  
PR 10-JUN-1998; 98US-US98826P.  
PR 11-JUN-1998; 98US-US98861P.  
PR 11-JUN-1998; 98US-US988663P.  
PR 11-JUN-1998; 98US-US98876P.  
PR 12-JUN-1998; 98US-US989090P.  
PR 12-JUN-1998; 98US-US989105P.  
PR 16-JUN-1998; 98US-US989512P.  
PR 16-JUN-1998; 98US-US989514P.  
PR 17-JUN-1998; 98US-US989538P.  
PR 17-JUN-1998; 98US-US989598P.  
PR 17-JUN-1998; 98US-US989598P.  
PR 18-JUN-1998; 98US-US989908P.  
PR 19-JUN-1998; 98US-US989952P.  
PR 22-JUN-1998; 98US-US990246P.  
PR 22-JUN-1998; 98US-US990252P.  
PR 22-JUN-1998; 98US-US990254P.  
PR 24-JUN-1998; 98US-US990429P.  
PR 24-JUN-1998; 98US-US990435P.  
PR 24-JUN-1998; 98US-US990444P.  
PR 24-JUN-1998; 98US-US990461P.  
PR 24-JUN-1998; 98US-US990535P.  
PR 24-JUN-1998; 98US-US990540P.  
PR 25-JUN-1998; 98US-US990676P.  
PR 25-JUN-1998; 98US-US990678P.  
PR 25-JUN-1998; 98US-US990688P.  
PR 25-JUN-1998; 98US-US990690P.  
PR 25-JUN-1998; 98US-US990694P.  
PR 25-JUN-1998; 98US-US990695P.  
PR 25-JUN-1998; 98US-US990696P.  
PR 26-JUN-1998; 98US-US990862P.  
PR 26-JUN-1998; 98US-US990863P.

Query Match 2.0%; Score 8; DB 24; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;

	Matches	8;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	81	SSGASTAT 88								
Db	39	SSGASTAT 46								
RESULT 59										
ABU57085	ABU57085 standard; Protein; 596 AA.									
ID	ABU57085									
XX	AC	ABU57085;								
XX	DT	04-APR-2003 (first entry)								
XX	DE	Human PRO polypeptide #155.								
XX	KW	Human; PRO; tumour necrosis factor-alpha; blood; cancer;								
KW	KW	chondrocyte cell; tumour; adrenal tumour; lung; colon; breast; prostate;								
KW	KW	kidney; rectum; cervix; liver; bone disorder; cartilage disorder;								
KW	KW	arthritis; sports injury; genetic disorder; antiarthritic; vulnery.								
XX	OS	Homo sapiens.								
XX	PN	US2003027280-A1.								
XX	PD	06-FEB-2003.								
XX	PF	20-JUN-2002; 2002US-0176993.								
XX	PR	16-SEP-1998; 98WO-US19330.								
PR	PR	07-OCT-1998; 98WO-US21141.								
PR	PR	01-DEC-1998; 98WO-US25108.								
PR	PR	08-MAR-1999; 99WO-US05028.								
PR	PR	14-MAY-1999; 99WO-US10733.								
PR	PR	02-JUN-1999; 99WO-US12252.								
PR	PR	01-SEP-1999; 99WO-US20111.								
PR	PR	15-SEP-1999; 99WO-US21090.								
PR	PR	01-DEC-1999; 99WO-US28301.								
PR	PR	02-DEC-1999; 99WO-US28551.								
PR	PR	30-DEC-1999; 99WO-US31274.								
PR	PR	05-JAN-2000; 2000WO-US00219.								
PR	PR	18-FEB-2000; 2000WO-US04341.								
PR	PR	18-FEB-2000; 2000WO-US04342.								
PR	PR	22-FEB-2000; 2000WO-US04414.								
PR	PR	24-FEB-2000; 2000WO-US05004.								
PR	PR	01-MAR-2000; 2000WO-US05601.								
PR	PR	02-MAR-2000; 2000WO-US05841.								
PR	PR	15-MAR-2000; 2000WO-US06884.								
PR	PR	30-MAR-2000; 2000WO-US08439.								
PR	PR	17-MAY-2000; 2000WO-US13705.								
PR	PR	22-MAY-2000; 2000WO-US14042.								
PR	PR	30-MAY-2000; 2000WO-US14941.								
PR	PR	02-JUN-2000; 2000WO-US15264.								
PR	PR	28-JUL-2000; 2000WO-US20710.								
PR	PR	24-AUG-2000; 2000WO-US23328.								
PR	PR	01-DEC-2000; 2000WO-US30952.								
PR	PR	20-DEC-2000; 2000WO-US32678.								
PR	PR	28-FEB-2001; 2001WO-US34956.								
PR	PR	01-JUN-2001; 2001WO-US06520.								
PR	PR	20-JUN-2001; 2001WO-US19692.								
PR	PR	29-JUN-2001; 2001WO-US21066.								
PR	PR	09-JUL-2001; 2001WO-US21735.								
PR	PR	29-AUG-2001; 2001WO-US27099.								
PR	PR	18-SEP-1997; 97US-059263P.								
PR	PR	18-SEP-1997; 97US-059266P.								
PR	PR	17-OCT-1997; 97US-062250P.								
PR	PR	21-OCT-1997; 97US-063486P.								
PR	PR	24-OCT-1997; 97US-063120P.								
PR	PR	24-OCT-1997; 97US-063121P.								
PR	PR	28-OCT-1997; 97US-063540P.								
PR	PR	28-OCT-1997; 97US-063541P.								

---

PR	28-OCT-1997;	97US-063544P.
PR	28-OCT-1997;	97US-063564P.
PR	29-OCT-1997;	97US-063734P.
PR	31-OCT-1997;	97US-063870P.
PR	31-OCT-1997;	97US-064103P.
PR	13-NOV-1997;	97US-065311P.
PR	21-NOV-1997;	97US-066120P.
PR	24-NOV-1997;	97US-066466P.
PR	11-DEC-1997;	97US-066772P.
PR	12-DEC-1997;	97US-069335P.
PR	17-DEC-1997;	97US-069425P.
PR	18-DEC-1997;	97US-069870P.
PR	10-MAR-1998;	97US-068017P.
PR	11-MAR-1998;	98US-077632P.
PR	11-MAR-1998;	98US-077649P.
PR	20-MAR-1998;	98US-078866P.
PR	20-MAR-1998;	98US-078939P.
PR	27-MAR-1998;	98US-079664P.
PR	27-MAR-1998;	98US-079786P.
PR	31-MAR-1998;	98US-080107P.
PR	31-MAR-1998;	98US-080194P.
PR	01-APR-1998;	98US-080327P.
PR	01-APR-1998;	98US-080333P.
PR	08-APR-1998;	98US-081049P.
PR	08-APR-1998;	98US-081070P.
PR	09-APR-1998;	98US-081195P.
PR	15-APR-1998;	98US-081838P.
PR	21-APR-1998;	98US-082568P.
PR	21-APR-1998;	98US-082569P.
PR	22-APR-1998;	98US-082704P.
PR	28-APR-1998;	98US-082797P.
PR	28-APR-1998;	98US-083322P.
PR	29-APR-1998;	98US-083495P.
PR	29-APR-1998;	98US-083496P.
PR	29-APR-1998;	98US-083499P.
PR	05-MAY-1998;	98US-083559P.
PR	06-MAY-1998;	98US-084366P.
PR	07-MAY-1998;	98US-084414P.
PR	07-MAY-1998;	98US-084639P.
PR	07-MAY-1998;	98US-084640P.
PR	07-MAY-1998;	98US-084643P.
PR	15-MAY-1998;	98US-085579P.
PR	15-MAY-1998;	98US-085580P.
PR	15-MAY-1998;	98US-085582P.
PR	18-MAY-1998;	98US-086023P.
PR	22-MAY-1998;	98US-086392P.
PR	22-MAY-1998;	98US-086486P.
PR	28-MAY-1998;	98US-087098P.
PR	28-MAY-1998;	98US-087208P.
PR	02-JUN-1998;	98US-087609P.
PR	02-JUN-1998;	98US-087759P.
PR	03-JUN-1998;	98US-087827P.
PR	04-JUN-1998;	98US-088025P.
PR	04-JUN-1998;	98US-088028P.
PR	04-JUN-1998;	98US-088029P.
PR	04-JUN-1998;	98US-088033P.
PR	04-JUN-1998;	98US-088326P.
PR	05-JUN-1998;	98US-088167P.
PR	05-JUN-1998;	98US-088202P.
PR	05-JUN-1998;	98US-088212P.
PR	05-JUN-1998;	98US-088217P.
PR	05-JUN-1998;	98US-088655P.
PR	10-JUN-1998;	98US-088722P.
PR	10-JUN-1998;	98US-088738P.
PR	10-JUN-1998;	98US-088740P.
PR	10-JUN-1998;	98US-088811P.
PR	10-JUN-1998;	98US-088824P.
PR	10-JUN-1998;	98US-088825P.
PR	10-JUN-1998;	98US-088826P.
PR	11-JUN-1998;	98US-088861P.
PR	11-JUN-1998;	98US-088863P.

PR 11-JUN-1998; 98US-088876P.  
PR 12-JUN-1998; 98US-089090P.  
PR 12-JUN-1998; 98US-089105P.  
PR 16-JUN-1998; 98US-089512P.  
PR 16-JUN-1998; 98US-089514P.  
PR 17-JUN-1998; 98US-089538P.  
PR 17-JUN-1998; 98US-089598P.  
PR 17-JUN-1998; 98US-089653P.  
PR 18-JUN-1998; 98US-089908P.  
PR 19-JUN-1998; 98US-089952P.  
PR 22-JUN-1998; 98US-090246P.  
PR 22-JUN-1998; 98US-090252P.  
PR 22-JUN-1998; 98US-090254P.  
PR 24-JUN-1998; 98US-090429P.  
PR 24-JUN-1998; 98US-090435P.  
PR 24-JUN-1998; 98US-090444P.  
PR 24-JUN-1998; 98US-090461P.  
PR 24-JUN-1998; 98US-090535P.  
PR 24-JUN-1998; 98US-090540P.  
PR 25-JUN-1998; 98US-090676P.  
PR 25-JUN-1998; 98US-090678P.  
PR 25-JUN-1998; 98US-090688P.  
PR 25-JUN-1998; 98US-090690P.  
PR 25-JUN-1998; 98US-090694P.  
PR 25-JUN-1998; 98US-090695P.  
PR 25-JUN-1998; 98US-090696P.  
PR 26-JUN-1998; 98US-090862P.  
PR 26-JUN-1998; 98US-090863P.  
PR 26-JUN-1998; 98US-091010P.  
PR 01-JUL-1998; 98US-091359P.  
PR 01-JUL-1998; 98US-091544P.  
PR 02-JUL-1998; 98US-091478P.  
PR 02-JUL-1998; 98US-091486P.  
PR 02-JUL-1998; 98US-091526P.  
PR 02-JUL-1998; 98US-091528P.  
PR 02-JUL-1998; 98US-091532P.  
PR 24-JUL-1998; 98US-094006P.  
PR 04-AUG-1998; 98US-095282P.  
PR 10-AUG-1998; 98US-095998P.  
PR 10-AUG-1998; 98US-096012P.  
PR 17-AUG-1998; 98US-096757P.  
PR 17-AUG-1998; 98US-096766P.  
PR 17-AUG-1998; 98US-096867P.  
PR 17-AUG-1998; 98US-096891P.  
PR 17-AUG-1998; 98US-096897P.  
PR 18-AUG-1998; 98US-096949P.  
PR 18-AUG-1998; 98US-096959P.  
PR 18-AUG-1998; 98US-097022P.  
PR 26-AUG-1998; 98US-097952P.  
PR 26-AUG-1998; 98US-097954P.  
PR 26-AUG-1998; 98US-097955P.  
PR 26-AUG-1998; 98US-097971P.  
PR 26-AUG-1998; 98US-097974P.  
PR 26-AUG-1998; 98US-098014P.  
PR 01-SEP-1998; 98US-098716P.  
PR 01-SEP-1998; 98US-098723P.  
PR 02-SEP-1998; 98US-098803P.  
PR 02-SEP-1998; 98US-098821P.  
PR 02-SEP-1998; 98US-098843P.  
PR 09-SEP-1998; 98US-099602P.  
PR 10-SEP-1998; 98US-099741P.  
PR 10-SEP-1998; 98US-099754P.  
PR 10-SEP-1998; 98US-099763P.  
PR 11-JUN-1998; 98US-088876P.  
PR 12-JUN-1998; 98US-089090P.  
PR 12-JUN-1998; 98US-089105P.  
PR 16-JUN-1998; 98US-089512P.  
PR 16-JUN-1998; 98US-089514P.  
PR 17-JUN-1998; 98US-089538P.  
PR 17-JUN-1998; 98US-089598P.  
PR 17-JUN-1998; 98US-089653P.  
PR 18-JUN-1998; 98US-089908P.  
PR 19-JUN-1998; 98US-089952P.  
PR 22-JUN-1998; 98US-090246P.  
PR 22-JUN-1998; 98US-090252P.  
PR 22-JUN-1998; 98US-090254P.  
PR 24-JUN-1998; 98US-090429P.  
PR 24-JUN-1998; 98US-090435P.  
PR 24-JUN-1998; 98US-090444P.  
PR 24-JUN-1998; 98US-090461P.  
PR 24-JUN-1998; 98US-090535P.  
PR 24-JUN-1998; 98US-090540P.  
PR 25-JUN-1998; 98US-090676P.  
PR 25-JUN-1998; 98US-090678P.  
PR 25-JUN-1998; 98US-090688P.  
PR 25-JUN-1998; 98US-090690P.  
PR 25-JUN-1998; 98US-090694P.  
PR 25-JUN-1998; 98US-090695P.  
PR 25-JUN-1998; 98US-090696P.  
PR 26-JUN-1998; 98US-090862P.  
PR 26-JUN-1998; 98US-090863P.  
PR 26-JUN-1998; 98US-091010P.  
PR 01-JUL-1998; 98US-091359P.  
PR 01-JUL-1998; 98US-091544P.  
PR 02-JUL-1998; 98US-091478P.  
PR 02-JUL-1998; 98US-091486P.  
PR 02-JUL-1998; 98US-091526P.  
PR 02-JUL-1998; 98US-091528P.  
PR 02-JUL-1998; 98US-091532P.  
PR 24-JUL-1998; 98US-094006P.  
PR 04-AUG-1998; 98US-095282P.  
PR 10-AUG-1998; 98US-095998P.  
PR 10-AUG-1998; 98US-096012P.  
PR 17-AUG-1998; 98US-096757P.  
PR 17-AUG-1998; 98US-096766P.  
PR 17-AUG-1998; 98US-096867P.  
PR 17-AUG-1998; 98US-096891P.  
PR 17-AUG-1998; 98US-096897P.  
PR 18-AUG-1998; 98US-096949P.  
PR 18-AUG-1998; 98US-096959P.  
PR 18-AUG-1998; 98US-097022P.  
PR 26-AUG-1998; 98US-097952P.  
PR 26-AUG-1998; 98US-097954P.  
PR 26-AUG-1998; 98US-097955P.  
PR 26-AUG-1998; 98US-097971P.  
PR 26-AUG-1998; 98US-097974P.  
PR 26-AUG-1998; 98US-098014P.  
PR 01-SEP-1998; 98US-098716P.  
PR 01-SEP-1998; 98US-098723P.  
PR 02-SEP-1998; 98US-098803P.  
PR 02-SEP-1998; 98US-098821P.  
PR 02-SEP-1998; 98US-098843P.  
PR 09-SEP-1998; 98US-099602P.  
PR 10-SEP-1998; 98US-099741P.  
PR 10-SEP-1998; 98US-099754P.  
PR 10-SEP-1998; 98US-099763P.  
Query Match  
Best Local Similarity 2.0%; Score 8; DB 24; Length 596;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Pred. No. 1.6e+02;  
QY 81 SSGASTAT 88  
DB 39 SSGASTAT 46

## RESULT 60

ABU10664

XX ABU10664 standard; Protein; 596 AA.

AC ABU10664;

XX 03-FEB-2003 (first entry)

XX Human secreted/transmembrane protein #155.

XX Human; secreted and transmembrane protein; blood;

XX tumour necrosis factor-alpha; chondrocyte cell proliferation;

XX colon tumour; breast tumour; prostate tumour; rectal tumour;

XX cervical tumour; liver tumour; bone disorder; cartilage disorder;

XX arthritis; sports injury.

OS Homo sapiens.

XX US2002127584-A1.

XX 12-SEP-2002.

XX 15-JAN-2002; 2002US-0052586.

XX 16-SEP-1998; 98WO-US19330.

XX 07-OCT-1998; 98WO-US21141.

XX 01-DEC-1998; 98WO-US25108.

XX 06-JAN-1999; 2000WO-US00219.

XX 08-MAR-1999; 99WO-US05028.

XX 14-MAY-1999; 99WO-US10733.

XX 02-JUN-1999; 99WO-US12252.

XX 01-SEP-1999; 99WO-US20111.

XX 15-SEP-1999; 99WO-US21030.

XX 01-DEC-1999; 99WO-US28301.

XX 30-DEC-1999; 99WO-US28551.

XX 18-FEB-2000; 2000WO-US04341.

XX 18-FEB-2000; 2000WO-US04342.

XX 22-FEB-2000; 2000WO-US04414.

XX 24-FEB-2000; 2000WO-US05004.

XX 01-MAR-2000; 2000WO-US05601.

XX 02-MAR-2000; 2000WO-US05841.

XX 15-MAR-2000; 2000WO-US06884.

XX 30-MAR-2000; 2000WO-US08439.

XX 17-MAY-2000; 2000WO-US13705.

XX 22-MAY-2000; 2000WO-US14042.

XX 30-MAY-2000; 2000WO-US14941.

XX 02-JUN-2000; 2000WO-US15264.

XX 28-JUL-2000; 2000WO-US20710.

XX 24-AUG-2000; 2000WO-US23328.

XX 08-NOV-2000; 2000WO-US30952.

XX 01-DEC-2000; 2000WO-US32678.

XX 20-DEC-2000; 2000WO-US34956.

XX 28-FEB-2001; 2001WO-US06520.

XX 01-JUN-2001; 2001WO-US17800.

XX 20-JUN-2001; 2001WO-US19632.

XX 29-JUN-2001; 2001WO-US21086.

XX 09-JUL-2001; 2001WO-US21735.

XX 29-AUG-2001; 2001WO-US27099.

XX 18-SEP-1997; 97US-059263P.

XX 18-SEP-1997; 97US-059266P.

XX 17-OCT-1997; 97US-062250P.

XX 21-OCT-1997; 97US-063486P.

XX 24-OCT-1997; 97US-063120P.

XX 24-OCT-1997; 97US-063121P.

XX 28-OCT-1997; 97US-063540P.

XX 28-OCT-1997; 97US-063541P.

XX 28-OCT-1997; 97US-063544P.

XX 28-OCT-1997; 97US-063564P.

XX 29-OCT-1997; 97US-063734P.

XX 31-OCT-1997; 97US-063870P.

XX 31-OCT-1997; 97US-064103P.

PR 13-NOV-1997; 97US-065311P.  
 PR 21-NOV-1997; 97US-066120P.  
 PR 24-NOV-1997; 97US-066466P.  
 PR 24-NOV-1997; 97US-066772P.  
 PR 11-DEC-1997; 97US-069335P.  
 PR 12-DEC-1997; 97US-069425P.  
 PR 17-DEC-1997; 97US-069870P.  
 PR 18-DEC-1997; 97US-068017P.  
 PR 10-MAR-1998; 98US-077450P.  
 PR 11-MAR-1998; 98US-077632P.  
 PR 11-MAR-1998; 98US-077649P.  
 PR 20-MAR-1998; 98US-078886P.  
 PR 20-MAR-1998; 98US-078939P.  
 PR 27-MAR-1998; 98US-079664P.  
 PR 31-MAR-1998; 98US-079786P.  
 PR 31-MAR-1998; 98US-080107P.  
 PR 31-MAR-1998; 98US-080194P.  
 PR 01-APR-1998; 98US-080327P.  
 PR 01-APR-1998; 98US-080333P.  
 PR 08-APR-1998; 98US-081049P.  
 PR 08-APR-1998; 98US-081070P.  
 PR 05-APR-1998; 98US-081195P.  
 PR 15-APR-1998; 98US-081838P.  
 PR 21-APR-1998; 98US-082568P.  
 PR 21-APR-1998; 98US-082569P.  
 PR 22-APR-1998; 98US-082704P.  
 PR 22-APR-1998; 98US-082797P.  
 PR 28-APR-1998; 98US-083322P.  
 PR 29-APR-1998; 98US-083455P.  
 PR 29-APR-1998; 98US-083496P.  
 PR 29-APR-1998; 98US-083499P.  
 PR 29-APR-1998; 98US-083559P.  
 PR 05-MAY-1998; 98US-084366P.  
 PR 06-MAY-1998; 98US-084414P.  
 PR 07-MAY-1998; 98US-084635P.  
 PR 07-MAY-1998; 98US-084640P.  
 PR 07-MAY-1998; 98US-084643P.  
 XX

(GETH ) GENENTECH INC.

PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;

WPI; 2003-066893/06.

N-PSDB; ABX16740.

XX Novel isolated PRO polypeptides e.g., PRO1079, PRO827, PRO791, PRO1131,  
 PT PRO1136, PRO1183, PRO1343, PRO1760, PRO1567 or PRO4333, useful for  
 PT stimulating release of tumor necrosis factor-alpha from human blood -

XX Claim 11; Fig 310; 701pp; English.

XX The invention relates to an isolated PRO polypeptide comprising at least  
 CC 80% sequence identity to the protein sequences appearing as ABU0510-  
 CC ABU0814 (including a version lacking its associated signal peptide, or  
 CC an isolated extracellular domain of a PRO polypeptide with or without  
 CC its associated signal peptide. Also included are the nucleic acids  
 CC encoding the PRO proteins (being secreted and transmembrane proteins)  
 CC appearing as ABX1636-ABX16590, PRO expression vectors, host cells,  
 CC chimeric PRO fusion proteins, an anti-PRO antibody and a PRO  
 CC derived oligonucleotide sequence. The PRO polypeptides are useful for  
 CC stimulating release of tumor necrosis factor-alpha from human blood.  
 CC The PRO polypeptide PRO6029 is useful for stimulating proliferation or  
 CC differentiation of chondrocyte cells. The PRO polypeptides as specified  
 CC in the specification and having differential expression in tumor cells,  
 CC are useful for detecting presence of tumor in a mammal (such as adrenal  
 CC tumour, lung tumour, colon tumour, breast tumour, prostate tumour, rectal  
 CC tumour, cervical tumour or liver tumour. The PRO polypeptide PRO6029 is  
 CC useful for treating various bone and/or cartilage disorders such as  
 CC arthritis, and sports injuries. The PRO polypeptides are useful for  
 CC screening compounds to identify ant/agonists. PRO nucleic acids  
 CC are useful as hybridisation probes, in chromosome and gene mapping,  
 CC in the generation of anti-sense RNA and DNA, for the preparation of PRO

CC polypeptides and for generating knock-out animals. The present  
 CC sequence represents a PRO polypeptide.

SQ Sequence 596 AA;

Query Match 2.0%; Score 8; DB 24; Length 596;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88

|||||

Db 39 SSGASTAT 46

RESULT 61

ABU05975

ID ABU05975 standard; Protein; 643 AA.

XX

AC ABU05975;

XX

DT 08-APR-2003 (first entry)

XX

DE M. tuberculosis and M. leprae marker protein #626.

XX

KW Mycobacterioses; survival; virulence; protective antigen; vaccine;  
 KW Mycobacterial disease; tuberculosis; leprosy.

OS

OS Mycobacterium tuberculosis.

OS

OS Mycobacterium leprae.

XX

PN WO200274903-A2.

XX

PD 26-SEP-2002.

XX

PF 22-FEB-2002; 2002WO-IB01973.

XX

PR 22-FEB-2001; 2001US-270123P.

XX

PA (INSP ) INST PASTEUR.

XX

PI Cole S;

XX

DR WPI; 2002-759885/82.

XX

PT Identifying and selecting genes for survival or virulence of  
 PT mycobacteria by a comparative genomic analysis of the sequences of  
 PT Mycobacterium tuberculosis and M. leprae -

XX

PS Claim 17; Page 844-845; 874pp; English.

XX

CC This invention relates to a novel method for identifying essential genes  
 CC for survival or virulence of mycobacteria species. The method comprises  
 CC aligning the genomic sequence of a first mycobacterium species on a  
 CC genomic sequence of a second mycobacterium species and selecting a  
 CC polynucleotide sequence that is highly conserved in both genomes with no  
 CC counterparts in other bacterial sequences and that corresponds  
 CC to an essential gene for the survival or virulence of mycobacterium  
 CC species. The method of the invention is useful for detecting M.

CC tuberculosis or M. leprae infection. The method reduces the number of  
 CC potential new targets and protective antigens for new drugs and vaccine  
 CC compositions to treat and prevent mycobacterial diseases, particularly  
 CC tuberculosis and leprosy. The present sequence represents a marker  
 CC protein from Mycobacterium tuberculosis and Mycobacterium leprae  
 CC identified using the method of the invention.

SQ Sequence 643 AA;

Query Match 2.0%; Score 8; DB 23; Length 643;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 LGQLAAAV 348

|||||

Db 29 LGQLAAAV 36

RESULT 62  
 ID AAB46707 standard; Protein; 892 AA.  
 XX  
 AC AAB46707;  
 XX  
 DT 12-APR-2001 (first entry)  
 XX  
 DE Herpes simplex virus DNA polymerase protein fragment SEQ ID NO 16.  
 XX  
 KW Genome; thermophilic enzyme; washing powder; bleaching.  
 XX  
 OS Herpesvirus.  
 XX  
 PN WO200075335-A2.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PF 02-JUN-2000; 2000WO-IB00893.  
 XX  
 PR 02-JUN-1999; 99US-0137120.  
 XX  
 PA (DECO-) DECODE GENETICS EHF.  
 XX  
 PI Hjoerleifsdottir S, Hreggvidsson GO, Fridjonsson OH, Aevarsson A;  
 PI Kristjansson JK;  
 XX  
 DR WPI; 2001-061727/07.  
 XX  
 PT Isolated nucleic acid from bacteriophage RM 378 encoding enzymes useful  
 PT in recombinant DNA technology -  
 XX  
 PS Disclosure; Figure 3A-P; 42pp; English.  
 XX  
 CC This invention describes a novel isolated nucleic molecule (I) comprising  
 CC the genome of bacteriophage RM 378. The invention also describes (1) an  
 CC isolated nucleic acid which encodes a polypeptide obtainable from  
 CC bacteriophage RM 378, or its active derivative or fragment; (2) an  
 CC isolated bacteriophage RM 378 (II); (3) a DNA construct (III) comprising  
 CC operatively linked to a regulatory sequence; (4) a host cell comprising  
 CC (III); and (5) an isolated polypeptide (IV) obtainable from (II), its  
 CC active derivative or fragment. Bacteriophage RM 378 is useful for  
 CC producing thermophilic enzymes useful in DNA research and commercial  
 CC settings (e.g. proteases and lipases used in washing powder, hydrolytic  
 CC enzymes used in bleaching). The isolated nucleic acid molecules and  
 CC vectors are useful in the manufacture of encoded polypeptide, as probes  
 CC for isolating homologous sequences (e.g. from other bacteriophage  
 CC species), as well as for detecting the presence of the bacteriophage in  
 CC a culture of host cells. The polypeptides can be used as a molecular  
 CC weight marker on SDS-PAGE gels or on molecular sieve gel filtration  
 CC columns. Because the host organism of the RM378 bacteriophage is a  
 CC thermophile, the enzymes and proteins of the RM378 bacteriophage are  
 CC significantly more thermostable than those of other (e.g. mesophilic)  
 CC bacteriophages, such as the T4 bacteriophage of *Escherichia coli*. The  
 CC enhanced stability of the enzymes and proteins of RM378 bacteriophage  
 CC allows their use under temperature conditions which would be prohibitive  
 CC for other enzymes, thus increasing the range of conditions which can be  
 CC employed not only in DNA research but also in commercial settings.  
 XX  
 SQ Sequence 892 AA;

Query Match 2.0%; Score 8; DB 22; Length 892;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 TLAQPAAP 102  
 |||||  
 Db 38 TLAQPAAP 45

RESULT 63  
 ID AAW78898 standard; Protein; 898 AA.  
 XX  
 AC AAW78898;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 21-DEC-1998 (first entry)  
 XX  
 DE Rat UNC-5 homologue UNC5H-1.  
 XX  
 KW UNC-5; UNC5H-1; rat; netrin receptor; cell migration;  
 KW axon guidance; diagnosis; therapy.  
 XX  
 OS Rattus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 580..594  
 FT /note= "peptide used to raise rabbit  
 FT polyclonal antisera"  
 XX  
 PN WO9837085-A1.  
 XX  
 PD 27-AUG-1998.  
 XX  
 PF 19-FEB-1998; 98WO-US03143.  
 XX  
 PR 19-FEB-1997; 97US-0808982.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Tessier-lavigne M, Leonardo ED, Hinck L, Masu M, Keinomasu K;  
 DR WPI; 1998-495364/42.  
 DR N-PSDB; AAV52940.  
 XX  
 PT Netrin-binding, vertebrate proteins - useful for diagnosis, therapy  
 PT and the biopharmaceutical industry  
 XX  
 PS Claim 1; Page 19-22; 32pp; English.  
 XX  
 CC UNC5H-1 and UNC5H-2 (see AAW78900) are rat homologues of  
 CC Caenorhabditis elegans UNC-5 protein. Their amino acid sequences  
 CC were deduced from isolated unc5h cDNA clones (see AAV52940 and  
 CC AAV52942) isolated from an E18 brain cDNA library. The predicted  
 CC proteins show similarity with UNC-5, possess 2 predicted Ig-like  
 CC domains and 2 predicted thrombospondin type-1 repeats, a predicted  
 CC membrane spanning region, and a large intracellular domain. They  
 CC are predicted to be involved in cell migration and axon guidance,  
 CC and are characterised as receptor proteins for netrins. Human  
 CC UNC5H-1 (see AAW78899) and UNC5H-2 (see AAW78901) proteins are also  
 CC claimed. Vertebrate UNC-5 proteins may be produced recombinantly  
 CC from transfected host cells. The invention also provides unc-5  
 CC hybridisation probes and primers, vertebrate UNC-5-specific binding  
 CC agents such as specific antibodies, and methods of making and using  
 CC the subject compositions in diagnosis (e.g. genetic hybridisation  
 CC screens for vertebrate unc-5 transcripts), therapy (e.g. gene  
 CC therapy to modulate vertebrate unc-5 gene expression) and in the  
 CC biopharmaceutical industry (e.g. as immunogens, reagents for  
 CC modulating cell guidance, reagents for screening chemical libraries  
 CC for lead pharmacological agents, etc.).  
 CC (Updated on 25-MAR-2003 to correct PI field.)  
 XX  
 SQ Sequence 898 AA;

Query Match 2.0%; Score 8; DB 19; Length 898;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 341 LGQLAAAV 348  
 |||||  
 Db 873 LGQLAAAV 880

## RESULT 64

AAU97900  
ID AAU97900 standard; Protein; 898 AA.  
XX  
AC AAU97900;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Rat netrin binding membrane receptor UNC5H-1 protein.  
XX  
KW Netrin binding membrane receptor; receptor; UNC5H-1;  
KW Rat; neurotrophic; neuroprotective; cytoskeletal; antiparkinsonian;  
KW cerebroprotective; cancer; central nervous system; CNS; stroke;  
KW Parkinson's disease; multiple sclerosis; Alzheimer's disease.  
XX  
OS Rattus sp.  
XX  
FH Key Location/Qualifiers  
FT Domain 152..223  
FT /note= "Immunoglobulin domain"  
FT Domain 247..294  
FT /note= "Thrombospondine type 1 domain"  
FT Domain 302..348  
FT /note= "Thrombospondine type 1 domain"  
FT Region 361..382  
FT /note= "Transmembrane region"  
FT Domain 495..598  
FT /note= "ZUS domain"  
FT Domain 817..897  
FT /note= "Death domain"  
XX  
PN WO200233080-A2.  
XX  
PD 25-APR-2002.  
XX  
PF 15-OCT-2001; 2001WO-EP11891.  
XX  
PR 16-OCT-2000; 2000US-240061P.  
XX  
PA (FARB ) BAYER AG.  
XX  
PI Koehler RH;  
XX  
DR WPI; 2002-463314/49.  
XX  
PT Novel human netrin binding membrane receptor polypeptide and  
PT polynucleotides for identifying modulating agents useful in treating  
PT diseases e.g. Parkinson's disease, multiple sclerosis, stroke,  
PT Alzheimer's disease  
XX  
PS Disclosure; Fig 3; 94pp; English.  
XX  
CC This invention relates to the DNA and protein sequences of a novel  
CC purified human netrin binding membrane receptor, UNC5H-1. The DNA  
CC sequence of the invention is useful as a probe for detecting a  
CC nucleic acid encoding the UNC5H-1 protein in a biological sample.  
CC The sequences of the invention are useful to screen for agents which  
CC decrease the activity of the UNC5H-1 protein. The sequences are also  
CC useful for screening agents which regulate (modulate) the activity of  
CC the protein of the invention. A pharmaceutical composition containing  
CC the protein of the invention or a reagent that modulates the activity  
CC of the UNC5H-1 protein may be useful for treating a UNC5H-1 dysfunction  
CC related disease such as cancer or a central nervous system (CNS)  
CC disorders (e.g. Parkinson's disease, multiple sclerosis, stroke and  
CC Alzheimer's disease). Fusion proteins comprising the UNC5H-1 protein are  
CC useful for generating antibodies and for in various assay systems, and  
CC the protein can be used as a bait protein in a two-hybrid assay or  
CC three-hybrid assay. The method of the invention is useful for detecting  
CC a coding sequence for the UNC5H-1 protein. The present sequence  
CC represents the Rat netrin binding membrane receptor UNC5H-1 protein of  
CC the invention.

## SQ Sequence 898 AA;

Query Match 2.0%; Score 8; DB 23; Length 898;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 341 LQQLAAAV 348  
DB 873 LQQLAAAV 880  
XX  
DE 14-FEB-2002 (first entry)  
XX  
DE Rat netrin receptor UNC5H1 (YSG7) polypeptide.  
XX  
KW YSG; YSG7; schizophrenia; chronic animal model; LCGU; netrin receptor;  
KW local cerebral glucose utilisation; phosphodiesterase 1-alpha; UNC5H1;  
KW calcium-independent alpha-latrotoxin receptor; CIRL; trkE; synapsin 1A;  
KW epithelial discoidin domain receptor 1; synapsin 1B; neuroleptic;  
KW tumour necrosis factor alpha; TNF-alpha; rat.  
XX  
OS Rattus sp.  
XX  
PN WO200175440-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 02-APR-2001; 2001WO-GB01486.  
XX  
PR 31-MAR-2000; 2000GB-0007880.  
XX  
PR 26-MAY-2000; 2000GB-0012768.  
XX  
PA (WELF-) WELFIDE CORP.  
XX  
PI Cochran S, Paterson G, Ohashi Y, Morris B, Pratt J;  
XX  
DR WPI; 2002-010813/01.  
XX  
DR N-PSDB; AAS16843.  
XX  
PT Novel chronic animal model of schizophrenia, useful for identifying  
PT anti-psychotic drugs and genes that are associated with schizophrenia  
XX  
PS Disclosure; Fig 8b; 79pp; English.  
XX  
CC The invention relates to YSG polynucleotide fragments for use in  
CC diagnosing and/or developing treatments for schizophrenia using chronic  
CC animal models. The polynucleotides and their encoded polypeptides are  
CC used for identification of compounds which modulate the expression of YSG  
CC molecules, leading to the manufacture of schizophrenia medicaments. The  
CC sequences can also be used for testing candidate compounds for any effect  
CC on the polypeptides. Anti-schizophrenic effects of a compound can be  
CC determined by measuring local cerebral glucose utilisation (LCGU) or  
CC comparing its expression level with that of a control group. The  
CC sequences are useful in the identification of genes associated with  
CC schizophrenic states and in the development of an antibody. The sequences  
CC of the invention include phosphodiesterase 1-alpha, calcium-independent  
CC alpha-latrotoxin receptors (CIRL)-1-263, epithelial discoidin domain  
CC receptor 1 (trkE), netrin receptor (UNC5H1), synapsins 1A and 1B and  
CC tumour necrosis factor (TNF) alpha. This sequence represents rat netrin  
CC receptor UNC5H1 (YSG7) polypeptide.  
XX  
SQ Sequence 898 AA;

Query Match 2.0%; Score 8; DB 23; Length 898;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 LGQLAAAV 348  
 |||||  
 Db 873 LGQLAAAV 880

## RESULT 66

AAU34320  
 ID AAU34320 standard; Protein; 2478 AA.

XX  
 AC AAU34320;

DT 14-FEB-2002 (first entry)

XX Staphylococcus aureus cellular proliferation protein #596.

XX Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX

OS Staphylococcus aureus.

XX  
 PN WO200170955-A2.

PD 27-SEP-2001.

XX  
 PF 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207277P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX

DR WPI: 2001-611495/70.

DR N-PSDB; AAS52179.

XX

PT New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX

PS Example 3; Seq ID No 5816; 51lpp; English.

XX

CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

Sequence 2478 AA;

SQ

Query Match 2.0%; Score 8; DB 22; Length 2478;

Best Local Similarity 100.0%; Pred. No. 6.6e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQPAPVA 104  
 |||||  
 Db 161 AQPAPVA 168

## RESULT 67

AAU37374  
 ID AAU37374 standard; Protein; 2478 AA.

XX  
 AC AAU37374;

DT 14-FEB-2002 (first entry)

XX Staphylococcus aureus cellular proliferation protein #1544.

XX Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX

OS Staphylococcus aureus.

XX  
 PN WO200170955-A2.

PD 27-SEP-2001.

XX  
 PF 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207277P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX

DR WPI: 2001-611495/70.

DR N-PSDB; AAS55233.

XX

PT New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX

PS Example 3; Seq ID No 12967; 51lpp; English.

XX

CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

Sequence 2478 AA;

SQ

Query Match 2.0%; Score 8; DB 22; Length 2478;

Best Local Similarity 100.0%; Pred. No. 6.6e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 97 AQAAPVA 104  
|||||  
Db 161 AQAAPVA 168

## RESULT 68

ABJ19002  
ID ABJ19002 standard; Protein; 2478 AA.  
XX  
AC ABJ19002;

XX 06-MAR-2003 (first entry)

XX Pathogen specific antigen related staphylococcal protein SEQ ID No 220.

XX Antibacterial; virucide; fungicide; protozoicide; cytostatic; anti-HIV;  
KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
KW autoimmune disease; HIV; hepatitis.

XX Staphylococcus sp.

XX WO200259148-A2.

XX 01-AUG-2002.

XX 21-JAN-2002; 2002WO-EP00546.

XX 26-JAN-2001; 2001AT-0000130.

XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.

PI Meinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W;  
PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;  
PI Tempelmaier B;

XX WPI; 2003-075410/07.

XX Identifying, isolating and producing hyperimmune serum-reactive  
PT antigens from a pathogen, for preparing vaccine or medicament for  
PT treating or preventing e.g. staphylococcal infections, comprises  
PT providing antibody preparation -

XX Example 7; Page 186; 252pp; English.

XX The invention relates to a novel method for identifying, isolating and  
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
CC allergen, a tissue or host prone to auto-immunity, where the antigens  
CC are used in a vaccine, comprises providing antibody preparation from a  
CC plasma pool of a type of animal, or individual sera with antibodies  
CC against the specific pathogen, tumour, allergen, tissue or host prone to  
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of  
CC the 62 sequences of 53-2261 amino acids fully defined in the  
CC specification, or their hyperimmune fragments are useful for the  
CC manufacture of a pharmaceutical preparation, particularly a vaccine  
CC against staphylococcal infections or colonisation against *S. aureus* or *S.*  
CC epidermidis. The preparation of antibodies is useful for the manufacture  
CC of a medicament for treating or preventing staphylococcal infections or  
CC colonisation against *S. aureus* or *S. epidermidis*. The antibody  
CC preparations may also be used for diagnostic and imaging purposes. Other  
CC conditions that can be treated include cancer, autoimmune diseases or  
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or  
CC protozoan pathogens. This sequence represents a staphylococcal protein  
CC relating to the method for identifying and producing pathogen specific  
CC antigens of the invention.

XX SQ Sequence 2478 AA;

Query Match 2.0%; Score 8; DB 24; Length 2478;  
Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQAAPVA 104  
|||||  
Db 161 AQAAPVA 168

## RESULT 69

ABG96596  
ID ABG96596 standard; Peptide; 9 AA.

XX  
AC ABG96596;

XX 16-DEC-2002 (first entry)

XX Human leukocyte antigen (HLA) B15 ligand #19.

XX Soluble human leukocyte antigen; HLA; sHLA; cell pharm;  
KW multimeric HLA complex; cytotoxic T lymphocytes; CTL; immune response;  
KW viral infection; cancer; autoimmune disease; vaccine development; MHC;  
KW major histocompatibility complex; diagnostic development;  
KW HLA class I polymorphism; HLA-B15 allotype; ligand.

XX Homo sapiens.

XX WO200262846-A2.

XX 15-AUG-2002.

XX 18-DEC-2001; 2001WO-US49744.

XX 18-DEC-2000; 2000US-256409P.

XX 18-DEC-2000; 2000US-256410P.

XX 10-OCT-2001; 2001US-0974366.

XX (HILD/) HILDEBRAND W H.  
XX (PRIL/) PRILLIMAN K R.

XX Hildebrand WH, Prilliman KR;

XX WPI; 2002-698563/75.

XX Producing soluble human leukocyte antigen (HLA) in cell pharm useful  
PT for studies of peptide loading for characterizing human immune  
PT responses involves using HLA allelic cDNA or genomic DNA as starting  
PT material -

XX Disclosure; Page 159; 300pp; English.

XX The invention describes a method of producing soluble human leukocyte  
CC antigen (HLA) molecules (sHLA) in cell pharm involving amplifying HLA  
CC allelic DNA by PCR using a locus specific primer to produce truncated a  
CC PCR product (PI), inserting PI into mammalian expression vector;  
CC electroporating the plasmid into a host cell; inoculating the cell pharm  
CC with the host cell such that cell pharm produces sHLA. A multimeric HLA  
CC complex (I) is useful for testing functionality of peptide ligands bound  
CC by at least two soluble HLA molecules. (I) can be tested for its ability  
CC to serve as ligands for cytotoxic T lymphocytes (CTLs) and induce immune  
CC responses in humans. (I) is useful for studying T cell responses to  
CC pathological conditions such as viral infections and cancer, and for  
CC modulating the human immune system to induce tolerance in autoimmune  
CC diseases. The individual secreted major histocompatibility complex (MHC)  
CC molecules produced are useful for studies of peptide loading (i.e., in  
CC vaccine development) and to the development of diagnostics with the  
CC secreted MHC molecules, naturally loaded peptides can be eluted from the  
CC MHC molecule and characterised. The secreted MHC molecules allow the  
CC assessment of structural and functional impact of HLA class I  
CC polymorphism. The molecules are also useful to generate ligands and hence  
CC ligand maps from the peptide pools extracted from series of distinct yet  
CC related class I HLA-B15 allotypes; compare the different ligand maps to  
CC identify potentially shared elements; and characterise the elements  
CC identified to positively or negatively validate the occurrence of  
CC overlapping ligands. The truncated version of (MHC) can be produced in  
CC mammalian or insect/bacterial cells such that milligram or greater  
CC quantities of an individual class I or class II molecule can be obtained.

CC This sequence represents a HLA (human leukocyte antigen) peptide  
 CC ligand.  
 XX  
 SQ Sequence 9 AA;  
 Query Match 1.7%; Score 7; DB 23; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 284 QFQALLQ 290  
 Db 2 QFQALLQ 8  
 RESULT 70  
 ABG96636  
 ID ABG96636 standard; Peptide; 9 AA.  
 XX  
 AC ABG96636;  
 XX  
 DT 16-DEC-2002 (first entry)  
 XX  
 DE Human leukocyte antigen (HLA) B15 ligand #59.  
 XX  
 KW Soluble human leukocyte antigen; HLA; sHLA; cell pharm;  
 KW multimeric HLA complex; cytotoxic T lymphocytes; CTL; immune response;  
 KW viral infection; cancer; autoimmune disease; vaccine development; MHC;  
 KW major histocompatibility complex; diagnostic development;  
 KW HLA class I polymorphism; HLA-B15 allotype; ligand.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200262846-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 18-DEC-2001; 2001WO-US49744.  
 XX  
 PR 18-DEC-2000; 2000US-256409P.  
 PR 18-DEC-2000; 2000US-256410P.  
 PR 10-OCT-2001; 2001US-0974366.  
 XX  
 PA (HILD/) HILDEBRAND W H.  
 PA (PRIL/) PRILLIMAN K R.  
 XX  
 PI Hildebrand WH, Prilliman KR;  
 XX  
 DR WPI; 2002-698563/75.  
 XX  
 PT Producing soluble human leukocyte antigen (HLA) in cell pharm useful  
 PT for studies of peptide loading for characterizing human immune  
 PT responses involves using HLA allelic cDNA or genomic DNA as starting  
 PT material -  
 XX  
 PS Disclosure; Page 161; 300pp; English.  
 XX  
 CC The invention describes a method of producing soluble human leukocyte  
 CC antigen (HLA) molecules (sHLA) in cell pharm involving amplifying HLA  
 CC allelic DNA by PCR using a locus specific primer to produce truncated a  
 CC PCR product (PI), inserting PI into mammalian expression vector;  
 CC electroporating the plasmid into a host cell; inoculating the cell pharm  
 CC with the host cell such that cell pharm produces sHLA. A multimeric HLA  
 CC complex (I) is useful for testing functionality of peptide ligands bound  
 CC by at least two soluble HLA molecules. (I) can be tested for its ability  
 CC to serve as ligands for cytotoxic T lymphocytes (CTLs) and induce immune  
 CC responses in humans. (I) is useful for studying T cell responses to  
 CC pathological conditions such as viral infections and cancer, and for  
 CC modulating the human immune system to induce tolerance in autoimmune  
 CC diseases. The individual secreted major histocompatibility complex (MHC)  
 CC molecules produced are useful for studies of peptide loading (i.e., in  
 CC vaccine development) and to the development of diagnostics. With the  
 CC secreted MHC molecules, naturally loaded peptides can be eluted from the  
 CC MHC molecule and characterised. The secreted MHC molecules allow the

CC assessment of structural and functional impact of HLA class I  
 CC polymorphism. The molecules are also useful to generate ligands and hence  
 CC ligand maps from the peptide pools extracted from series of distinct yet  
 CC related class I HLA-B15 allotypes; compare the different ligand maps to  
 CC identify potentially shared elements; and characterise the elements  
 CC identified to positively or negatively validate the occurrence of  
 CC overlapping ligands. The truncated version of (MHC) can be produced in  
 CC mammalian or insect/bacterial cells such that milligram or greater  
 CC quantities of an individual class I or class II molecule can be obtained.  
 CC This sequence represents a HLA (human leukocyte antigen) peptide  
 CC ligand.  
 XX  
 SQ Sequence 9 AA;  
 Query Match 1.7%; Score 7; DB 23; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 284 QFQALLQ 290  
 Db 2 QFQALLQ 8  
 RESULT 71  
 ABG97255  
 ID ABG97255 standard; Peptide; 9 AA.  
 XX  
 AC ABG97255;  
 XX  
 DT 16-DEC-2002 (first entry)  
 XX  
 DE Human leukocyte antigen (HLA) B15 ligand #584.  
 XX  
 KW Soluble human leukocyte antigen; HLA; sHLA; cell pharm;  
 KW multimeric HLA complex; cytotoxic T lymphocytes; CTL; immune response;  
 KW viral infection; cancer; autoimmune disease; vaccine development; MHC;  
 KW major histocompatibility complex; diagnostic development;  
 KW HLA class I polymorphism; HLA-B15 allotype; ligand.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200262846-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 18-DEC-2001; 2001WO-US49744.  
 XX  
 PR 18-DEC-2000; 2000US-256409P.  
 PR 18-DEC-2000; 2000US-256410P.  
 PR 10-OCT-2001; 2001US-0974366.  
 XX  
 PA (HILD/) HILDEBRAND W H.  
 PA (PRIL/) PRILLIMAN K R.  
 XX  
 PI Hildebrand WH, Prilliman KR;  
 XX  
 DR WPI; 2002-698563/75.  
 XX  
 PT Producing soluble human leukocyte antigen (HLA) in cell pharm useful  
 PT for studies of peptide loading for characterizing human immune  
 PT responses involves using HLA allelic cDNA or genomic DNA as starting  
 PT material -  
 XX  
 PS Disclosure; Page 188; 300pp; English.  
 XX  
 CC The invention describes a method of producing soluble human leukocyte  
 CC antigen (HLA) molecules (sHLA) in cell pharm involving amplifying HLA  
 CC allelic DNA by PCR using a locus specific primer to produce truncated a  
 CC PCR product (PI), inserting PI into mammalian expression vector;  
 CC electroporating the plasmid into a host cell; inoculating the cell pharm  
 CC with the host cell such that cell pharm produces sHLA. A multimeric HLA  
 CC complex (I) is useful for testing functionality of peptide ligands bound  
 CC by at least two soluble HLA molecules. (I) can be tested for its ability

CC to serve as ligands for cytotoxic T lymphocytes (CTLs) and induce immune  
 CC responses in humans. (I) is useful for studying T cell responses to  
 CC pathological conditions such as viral infections and cancer, and for  
 CC modulating the human immune system to induce tolerance in autoimmune  
 CC diseases. The individual secreted major histocompatibility complex (MHC)  
 CC molecules produced are useful for studies of peptide loading (i.e., in  
 CC vaccine development) and to the development of diagnostics. With the  
 CC secreted MHC molecules, naturally loaded peptides can be eluted from the  
 CC MHC molecule and characterised. The secreted MHC molecules allow the  
 CC assessment of structural and functional impact of HLA class I  
 CC polymorphism. The molecules are also useful to generate ligands and hence  
 CC ligand maps from the peptide pools extracted from series of distinct yet  
 CC related class I HLA-B\*57 allotypes; compare the different ligand maps to  
 CC identify potentially shared elements; and characterise the elements  
 CC identified to positively or negatively validate the occurrence of  
 CC overlapping ligands. The truncated version of (MHC) can be produced in  
 CC mammalian or insect/bacterial cells such that milligram or greater  
 CC quantities of an individual class I or class II molecule can be obtained.  
 CC This sequence represents a HLA (human leukocyte antigen) peptide  
 CC ligand.

XX Sequence 9 AA;  
 SQ Query Match 1.7%; Score 7; DB 23; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 284 QFOALLQ 290  
 Db 2 QFOALLQ 8  
 RESULT 72  
 AAG94101  
 ID AAG94101 standard; Peptide; 10 AA.  
 AC AAG94101;  
 XX 18-SEP-2001 (first entry)  
 DE Human complementary peptide, SEQ ID NO: 295.  
 XX Human complementary peptide; ligand; drug discovery; drug design.  
 KW Human; complementary peptide; ligand; drug discovery; drug design.  
 XX Homo sapiens.  
 OS WO200142277-A2.  
 XX 14-JUN-2001.  
 PF 13-DEC-2000; 2000WO-GB04776.  
 XX 13-DEC-1999; 99GB-0029464.  
 PA (PROT-) PROTEOM LTD.  
 XX Roberts GW, Heal JR;  
 PI WPI; 2001-408419/43.  
 DR A set of peptide ligands consisting of specific complementary peptides  
 XX to proteins encoded by genes of the human genome, useful in an assay  
 PT for screening and identifying of one or more novel peptides which are  
 PT drug candidates or pro-drugs -  
 XX Example 4; Page 82; 646pp; English.  
 PS The invention relates to a set of complementary peptide ligands  
 XX generated from the human genome. The complementary peptides  
 CC interact with their relevant target proteins encoded in the human  
 CC genome. They can be used as reagents in drug discovery and as lead  
 CC ligands to facilitate drug design and development. The present  
 CC sequence is a complementary peptide provided in the specification.

XX Sequence 10 AA;  
 SQ Query Match 1.7%; Score 7; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 105 PAASVAR 111  
 Db 1 PAASVAR 7  
 RESULT 73  
 ABG02484  
 ID ABG02484 standard; Protein; 52 AA.  
 XX ABG02484;  
 AC 13-FEB-2002 (first entry)  
 DT Novel human diagnostic protein #2475.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 XX 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR N-PSDB; AAS66671.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 20; SEQ ID No 32843; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (III). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 52 AA;  
 SQ

Query Match 1.7%; Score 7; DB 22; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 PQAPAA 136  
 |||||  
 Db 3 PQAPAA 9

## RESULT 74

ABG57158  
 ID ABG57158 standard; Peptide; 60 AA.

XX AC ABG57158;  
 XX XX

DT 25-FEB-2003 (first entry)

DE Human liver peptide, SEQ ID No 35806.

XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 KW hypercholesterolaemia; coronary heart disease.  
 XX OS Homo sapiens.

XX WO200157273-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00664.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488898/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analysing gene expression in human adult liver -

XX Claim 27; SEQ ID No 35806; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SEN) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult  
 CC liver. (I) may be used for predicting, measuring and displaying gene  
 CC expression in samples derived from human adult liver. The genes  
 CC identified may be involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
 CC is associated with coronary heart disease. ABG47348-ABG59930 represent  
 CC human liver single exon encoded peptides of the invention.  
 CC Note: The sequence information for this patent does not appear in the  
 CC printed specification but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 60 AA;

Query Match 1.7%; Score 7; DB 22; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 PVAPAA 108  
 |||||

Db 42 PVAPAA 48

## RESULT 75

ABB41719  
 ID ABB41719 standard; Peptide; 60 AA.

XX AC ABB41719;  
 XX XX

DT 04-FEB-2002 (first entry)

DE Peptide #9225 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for  
 PT analysing gene expression in human fetal liver -  
 PT Claim 27; SEQ ID NO 34354; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC foetal liver. The present sequence is a peptide encoded by a single exon  
 CC nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 60 AA;

Query Match 1.7%; Score 7; DB 22; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 PVAPAA 108  
 |||||

Db 42 PVAPAA 48

Search completed: December 17, 2003, 06:33:18  
 Job time : 42.2471 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:35:27 ; Search time 34.0556 Seconds  
(without alignments)  
2220.991 Million cell updates/sec

Title: US-09-805-550-2  
Perfect score: 405  
Sequence: 1 MKNVTKLTGTFEIRASPD.....EELTANYLLDHGHEFPDQQQ 405

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 696363 seqs, 186758610 residues

Word size : 0

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Published Applications AA\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	405	100.0	405	9	US-09-805-550-2
2	13	3.2	368	9	US-09-805-550-4
3	9	2.2	3069	10	US-09-712-363-246
4	8	2.0	186	9	US-09-729-674-30
5	8	2.0	227	15	US-10-156-761-11941
6	8	2.0	254	9	US-09-815-242-13899
7	8	2.0	540	8	US-08-781-986A-5225
8	8	2.0	540	8	US-08-781-986A-5242
9	8	2.0	596	11	US-09-946-374-243
10	8	2.0	596	12	US-10-015-387A-243
11	8	2.0	596	12	US-10-063-735-100
12	8	2.0	596	12	US-10-006-130A-243
13	8	2.0	596	12	US-10-199-672-310
14	8	2.0	596	12	US-10-006-172A-243
15	8	2.0	596	12	US-10-187-749-310
16	8	2.0	596	12	US-10-063-582-100
17	8	2.0	596	12	US-10-063-581-100
18	8	2.0	596	12	US-10-063-582-100
19	8	2.0	596	12	US-10-063-582-100
20	8	2.0	596	12	US-10-063-582-100
21	8	2.0	596	12	US-10-063-582-100
22	8	2.0	596	12	US-10-063-582-100
23	8	2.0	596	12	US-10-063-582-100
24	8	2.0	596	12	US-10-063-582-100
25	8	2.0	596	12	US-10-063-582-100
26	8	2.0	596	12	US-10-063-582-100
27	8	2.0	596	12	US-10-063-582-100
28	8	2.0	596	12	US-10-063-582-100
29	8	2.0	596	12	US-10-063-582-100
30	8	2.0	596	12	US-10-063-582-100
31	8	2.0	596	12	US-10-063-582-100
32	8	2.0	596	12	US-10-063-582-100
33	8	2.0	596	12	US-10-063-582-100
34	8	2.0	596	12	US-10-063-582-100
35	8	2.0	596	12	US-10-063-582-100
36	8	2.0	596	12	US-10-063-582-100
37	8	2.0	596	12	US-10-063-582-100
38	8	2.0	596	12	US-10-063-582-100
39	8	2.0	596	12	US-10-063-582-100
40	8	2.0	596	12	US-10-063-582-100
41	8	2.0	596	12	US-10-063-582-100
42	8	2.0	596	12	US-10-063-582-100
43	8	2.0	596	12	US-10-063-582-100
44	8	2.0	596	12	US-10-063-582-100
45	8	2.0	596	12	US-10-063-582-100
46	8	2.0	596	12	US-10-063-582-100
47	8	2.0	596	12	US-10-063-582-100
48	8	2.0	596	12	US-10-063-582-100
49	8	2.0	596	12	US-10-063-582-100
50	8	2.0	596	12	US-10-063-582-100
51	8	2.0	596	12	US-10-063-582-100
52	8	2.0	596	12	US-10-063-582-100
53	8	2.0	596	12	US-10-063-582-100
54	8	2.0	596	12	US-10-063-582-100
55	8	2.0	596	12	US-10-063-582-100
56	8	2.0	596	12	US-10-063-582-100
57	8	2.0	596	12	US-10-063-582-100
58	8	2.0	596	12	US-10-063-582-100
59	8	2.0	596	12	US-10-063-582-100
60	8	2.0	596	12	US-10-063-582-100
61	8	2.0	596	12	US-10-063-582-100
62	8	2.0	596	12	US-10-063-582-100
63	8	2.0	596	12	US-10-063-582-100
64	8	2.0	596	12	US-10-063-582-100
65	8	2.0	596	12	US-10-063-582-100
66	8	2.0	596	12	US-10-063-582-100
67	8	2.0	596	12	US-10-063-582-100
68	8	2.0	596	12	US-10-063-582-100
69	8	2.0	596	12	US-10-063-582-100
70	8	2.0	596	12	US-10-063-582-100
71	8	2.0	596	12	US-10-063-582-100
72	8	2.0	596	12	US-10-063-582-100
73	8	2.0	596	12	US-10-063-582-100
74	8	2.0	596	12	US-10-063-582-100
75	8	2.0	596	12	US-10-063-582-100
76	8	2.0	596	12	US-10-063-582-100
77	8	2.0	596	12	US-10-063-582-100
78	8	2.0	596	12	US-10-063-582-100
79	8	2.0	596	12	US-10-063-582-100
80	8	2.0	596	12	US-10-063-582-100
81	8	2.0	596	12	US-10-063-582-100
82	8	2.0	596	12	US-10-063-582-100
83	8	2.0	596	12	US-10-063-582-100
84	8	2.0	596	12	US-10-063-582-100
85	8	2.0	596	12	US-10-063-582-100
86	8	2.0	596	12	US-10-063-582-100
87	8	2.0	596	12	US-10-063-582-100
88	8	2.0	596	12	US-10-063-582-100
89	8	2.0	596	12	US-10-063-582-100
90	8	2.0	596	12	US-10-063-582-100
91	8	2.0	596	12	US-10-063-582-100
92	8	2.0	596	12	US-10-063-582-100
93	8	2.0	596	12	US-10-063-582-100
94	8	2.0	596	12	US-10-063-582-100
95	8	2.0	596	12	US-10-063-582-100
96	8	2.0	596	12	US-10-063-582-100
97	8	2.0	596	12	US-10-063-582-100
98	8	2.0	596	12	US-10-063-582-100
99	8	2.0	596	12	US-10-063-582-100
100	8	2.0	596	12	US-10-063-582-100

```
89      8      2.0      596      12      US-10-063-583-100      Sequence 100, App
90      8      2.0      596      12      US-10-063-584-100      Sequence 100, App
91      8      2.0      596      12      US-10-063-587-100      Sequence 100, App
92      8      2.0      596      12      US-10-063-589-100      Sequence 100, App
93      8      2.0      596      12      US-10-063-591-100      Sequence 100, App
94      8      2.0      596      12      US-10-063-592-100      Sequence 100, App
95      8      2.0      596      12      US-10-063-593-100      Sequence 100, App
96      8      2.0      596      12      US-10-063-596-100      Sequence 100, App
97      8      2.0      596      12      US-10-063-597-100      Sequence 100, App
98      8      2.0      596      12      US-10-063-600-100      Sequence 100, App
99      8      2.0      596      12      US-10-063-602-100      Sequence 100, App
100     8      2.0      596      12      US-10-063-604-100      Sequence 100, App
```

## ALIGNMENTS

```
RESULT 1
US-09-805-550-2
; Sequence 2, Application US/09805550
; Patent No. US20020026045A1
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Rad23 Genes and Uses Thereof
; FILE REFERENCE: 0964D
; CURRENT APPLICATION NUMBER: US/09/805,550
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 09/413,574
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: 60/109,728
; PRIOR FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Zea mays
US-09-805-550-2

Query Match      100.0%; Score 405; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1      M K L N V K T L K G T N F E I E A S P D A S V A D V K R I I E T T Q G S T Y R A D Q Q M L I Y Q G K I L K D E T T L E      60
Db      1      M K L N V K T L K G T N F E I E A S P D A S V A D V K R I I E T T Q G S T Y R A D Q Q M L I Y Q G K I L K D E T T L E      60

Qy      61      S N G V A E N S F L V I M L S K A K A S S G A S T A T T A K A P A T L A Q P A A P A P A A S V A R T P T Q A P V A T      120
Db      61      S N G V A E N S F L V I M L S K A K A S S G A S T A T T A K A P A T L A Q P A A P A P A A S V A R T P T Q A P V A T      120

Qy      121     A E T A P S V Q P Q A P A A T V A A T D D A D V Y S Q A A S N L V F G N N L E Q T I Q Q I L D M G G G T W E R D T V      180
Db      121     A E T A P S V Q P Q A P A A T V A A T D D A D V Y S Q A A S N L V F G N N L E Q T I Q Q I L D M G G G T W E R D T V      180

Qy      181     V R A L P A A Y N N P E R A I D Y L I G S I P E N V E A Q P A R A P A A C Q Q T N Q Q A S P A Q P A V A L P V Q P S      240
Db      181     V R A L P A A Y N N P E R A I D Y L I G S I P E N V E A Q P A R A P A A C Q Q T N Q Q A S P A Q P A V A L P V Q P S      240

Qy      241     P A S A G P N A N P L N L F P Q G V P S G S N F G V P G A G S G A L D A L R O L P Q F Q A L L Q L V Q A N P Q I L Q      300
Db      241     P A S A G P N A N P L N L F P Q G V P S G S N F G V P G A G S G A L D A L R O L P Q F Q A L L Q L V Q A N P Q I L Q      300

Qy      301     P M L Q E L G K Q N P Q I L R L I Q E N Q A E F L R L V N E S P E G P G G N I L G Q L A A A A V P Q T L T V T P E E R E      360
Db      301     P M L Q E L G K Q N P Q I L R L I Q E N Q A E F L R L V N E S P E G P G G N I L G Q L A A A A V P Q T L T V T P E E R E      360

Qy      361     A I Q R L E G H G F N R E L V L E V F F A C N K D E E L T A N Y L L D H G H E F D D Q Q      405
Db      361     A I Q R L E G H G F N R E L V L E V F F A C N K D E E L T A N Y L L D H G H E F D D Q Q      405
```

```
RESULT 2
US-09-805-550-4
; Sequence 4, Application US/09805550
; Patent No. US20020026045A1
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Rad23 Genes and Uses Thereof
; FILE REFERENCE: 0964D
; CURRENT APPLICATION NUMBER: US/09/805,550
; CURRENT FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 09/413,574
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: 60/109,728
; PRIOR FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-09-805-550-4

Query Match      3.2%; Score 13; DB 9; Length 368;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      182     R A L R A A Y N N P E R A      194
Db      167     R A L R A A Y N N P E R A      179
```

```
RESULT 3
US-09-712-363-246
; Sequence 246, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 246
; LENGTH: 3069
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-246

Query Match      2.2%; Score 9; DB 10; Length 3069;
Best Local Similarity 100.0%; Pred. No. 64;
```

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 AAPVAPAA 108  
|||||  
Db 1754 AAPVAPAA 1762

## RESULT 4

US-09-729-674-30  
; Sequence 30, Application US/09729674  
; Patent No. US2001003935A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallie, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steininger II, Robert J.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Pechtel, Kim  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: 6055-64X  
; CURRENT APPLICATION NUMBER: US/09/729,674  
; CURRENT FILING DATE: 2000-12-04  
; PRIOR APPLICATION NUMBER: 09/539,330  
; PRIOR FILING DATE: 2000-03-30  
; NUMBER OF SEQ ID NOS: 283  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 30  
; LENGTH: 186  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-729-674-30

Query Match 2.0%; Score 8; DB 9; Length 186;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
|||||  
Db 39 SSGASTAT 46

## RESULT 5

US-10-156-761-11941  
; Sequence 11941, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 11941  
; LENGTH: 227  
; TYPE: PRT

; ORGANISM: Streptomyces avermitilis  
US-10-156-761-11941

Query Match 2.0%; Score 8; DB 15; Length 227;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 AAPVAPAA 107  
|||||  
Db 30 AAPVAPAA 37

## RESULT 6

US-09-815-242-13899  
; Sequence 13899, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13899  
; LENGTH: 254  
; TYPE: PRT  
; ORGANISM: Salmonella typhi  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(254)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-815-242-13899

Query Match 2.0%; Score 8; DB 9; Length 254;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 REAIQRL 366  
|||||  
Db 52 REAIQRL 59

## RESULT 7

US-08-781-986A-5225  
; Sequence 5225, Application US/08781986A  
; Publication No. US20030054436A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 5225:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 540 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-781-986A-5225

Query Match 2.0%; Score 8; DB 8; Length 540;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQAAPVA 104  
|||  
DB 175 AQAAPVA 182

RESULT 8  
US-08-781-986A-5242  
; Sequence 5242, Application US/08781986A  
; Publication No. US20030054436A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob

; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 5242:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 540 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-781-986A-5242

Query Match 2.0%; Score 8; DB 8; Length 540;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQAAPVA 104  
|||  
DB 175 AQAAPVA 182

RESULT 9  
US-09-946-374-243  
; Sequence 243, Application US/09946374  
; Publication No. US20030073129A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: P2830FIC1  
; CURRENT APPLICATION NUMBER: US/09/946,374  
; CURRENT FILING DATE: 2001-09-04  
; PRIOR APPLICATION NUMBER: 60/098716  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/098723  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/098749  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/098750  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/098803  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/098821  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/098843  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/099536  
; PRIOR FILING DATE: 1998-09-09  
; PRIOR APPLICATION NUMBER: 60/099596  
; PRIOR FILING DATE: 1998-09-09  
; PRIOR APPLICATION NUMBER: 60/099598



PRIOR APPLICATION NUMBER: 60/101479	PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738	PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101741	PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743	PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101915	PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916	PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207	PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240	PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307	PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330	PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331	PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484	PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487	PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102570	PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102571	PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102684	PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102687	PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102965	PRIOR FILING DATE: 1998-10-02
PRIOR APPLICATION NUMBER: 60/103258	PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103314	PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103315	PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103328	PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103395	PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103396	PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103401	PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103449	PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103633	PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103678	PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103679	PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987	PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103711	PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/104257	PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987	PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105000	PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002	PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105104	PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169	PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266	

; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: 60/105693  
; PRIOR FILING DATE: 1998-10-26  
; PRIOR APPLICATION NUMBER: 60/105694  
; PRIOR FILING DATE: 1998-10-26  
; PRIOR APPLICATION NUMBER: 60/105807

Query Match 2.0%; Score 8; DB 11; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 10

US-10-015-387A-243  
; Sequence 243, Application US/10015387A  
; Publication No. US20030135034A1

## GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan I.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; FILE REFERENCE: P2830P1C54

; CURRENT APPLICATION NUMBER: US/10/015,387A

; PRIOR FILING DATE: 2001-12-12

; PRIOR APPLICATION removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 477

; SEQ ID NO 243

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-015-387A-243

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 11

US-10-063-735-100  
; Sequence 100, Application US/10063735  
; Publication No. US20030138882A1

## GENERAL INFORMATION:

; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P2830P1C54

; CURRENT APPLICATION NUMBER: US/10/015,387A

; PRIOR FILING DATE: 2001-12-12

; PRIOR APPLICATION removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 477

; SEQ ID NO 243

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-015-387A-243

; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063,735  
; CURRENT FILING DATE: 2002-05-08

; PRIOR APPLICATION removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 170

; SEQ ID NO 100

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-063-735-100

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 12

US-10-006-130A-243  
; Sequence 243, Application US/10006130A  
; Publication No. US20030148375A1

## GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan I.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; FILE REFERENCE: P2830P1C7

; CURRENT APPLICATION NUMBER: US/10/006,130A

; PRIOR FILING DATE: 2002-03-19

; PRIOR APPLICATION removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 477

; SEQ ID NO 243

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-006-130A-243

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 13

US-10-199-672-310  
; Sequence 310, Application US/10199672  
; Publication No. US20030148442A1

## GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James

```

; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC1
; CURRENT APPLICATION NUMBER: US/10/199,672
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-199-672-310

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SSGASTAT 88
DB      39 SSGASTAT 46
|||||||

RESULT 14
US-10-006-172A-243
; Sequence 243, Application US/10006172A
; Publication No. US20030153000A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC11
; CURRENT APPLICATION NUMBER: US/10/006,172A
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099602
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099642
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099754
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099792
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099808
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099815
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100388
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100584
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100661
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100662
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100664
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100683
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100684
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100710
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100711
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100849
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/100930
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101014
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/101068
; PRIOR FILING DATE: 1998-09-18

```

; PRIOR APPLICATION NUMBER: 60/101071  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: 60/101279  
; PRIOR FILING DATE: 1998-09-22  
; PRIOR APPLICATION NUMBER: 60/101471  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101472  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101474  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101475  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101476  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101477  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101479  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101738  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101741  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101743  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101915  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101916  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/102207  
; PRIOR FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: 60/102240  
; PRIOR FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: 60/102307  
; PRIOR FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: 60/102330  
; PRIOR FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: 60/102331  
; PRIOR FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: 60/102484  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 60/102487  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 60/102570  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 60/102571  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 60/102684  
; PRIOR FILING DATE: 1998-10-01  
; PRIOR APPLICATION NUMBER: 60/102687  
; PRIOR FILING DATE: 1998-10-01  
; PRIOR APPLICATION NUMBER: 60/102965  
; PRIOR FILING DATE: 1998-10-02  
; PRIOR APPLICATION NUMBER: 60/103258  
; PRIOR FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: 60/103314  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103315  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103328  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103395  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103396  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103401  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103449  
; PRIOR FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: 60/103633  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/103678  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/103679

; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/103711  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/104257  
; PRIOR FILING DATE: 1998-10-14  
; PRIOR APPLICATION NUMBER: 60/104987  
; PRIOR FILING DATE: 1998-10-20  
; PRIOR APPLICATION NUMBER: 60/105000  
; PRIOR FILING DATE: 1998-10-20  
; PRIOR APPLICATION NUMBER: 60/105002  
; PRIOR FILING DATE: 1998-10-20  
; PRIOR APPLICATION NUMBER: 60/105104  
; PRIOR FILING DATE: 1998-10-21  
; PRIOR APPLICATION NUMBER: 60/105169  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: 60/105266  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: 60/105693  
; PRIOR FILING DATE: 1998-10-26  
; PRIOR APPLICATION NUMBER: 60/105694  
; PRIOR FILING DATE: 1998-10-26  
; PRIOR APPLICATION NUMBER: 60/105807  
; PRIOR FILING DATE: 1998-10-27  
; PRIOR APPLICATION NUMBER: 60/105881  
; PRIOR FILING DATE: 1998-10-27  
; PRIOR APPLICATION NUMBER: 60/105882  
; PRIOR FILING DATE: 1998-10-27  
; PRIOR APPLICATION NUMBER: 60/106023  
; PRIOR FILING DATE: 1998-10-28

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
|||||  
Db 39 SSGASTAT 46

## RESULT 15

US-10-187-749-310  
; Sequence 310, Application US/10187749  
; Publication No. US20030153036A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: F3430RIC1  
; CURRENT APPLICATION NUMBER: US/10/187,749  
; CURRENT FILING DATE: 2002-07-01  
; PRIOR APPLICATION NUMBER: US/10/052,586  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486

/ PRIOR FILING DATE: 1997-10-21  
/ PRIOR APPLICATION NUMBER: 60/063540  
/ PRIOR FILING DATE: 1997-10-28  
/ PRIOR APPLICATION NUMBER: 60/063541  
/ PRIOR FILING DATE: 1997-10-28  
/ PRIOR APPLICATION NUMBER: 60/063544  
/ PRIOR FILING DATE: 1997-10-28  
/ Remaining Prior Application data removed - See File Wrapper or PALM.  
/ NUMBER OF SEQ ID NOS: 612  
/ SEQ ID NO 310  
/ LENGTH: 596  
/ TYPE: PRT  
/ ORGANISM: Homo Sapien  
US-10-187-749-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 16

US-10-194-457-310  
/ Sequence 310, Application US/10194457  
/ Publication No. US20030153037A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Baker, Kevin P.  
/ APPLICANT: Chen, Jian  
/ APPLICANT: Desnoyers, Luc  
/ APPLICANT: Goddard, Audrey  
/ APPLICANT: Godowski, Paul J.  
/ APPLICANT: Gurney, Austin L.  
/ APPLICANT: Pan, James  
/ APPLICANT: Smith, Victoria  
/ APPLICANT: Watanabe, Colin K.  
/ APPLICANT: Wood, William I.  
/ APPLICANT: Zhang, Zemin  
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430RIC296

CURRENT FILING DATE: 2002-07-11

PRIOR APPLICATION NUMBER: 10/052586

PRIOR FILING DATE: 2002-01-15

PRIOR APPLICATION NUMBER: 60/059263

PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/059266

PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/062250

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/063120

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063121

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063486

PRIOR FILING DATE: 1997-10-21

PRIOR APPLICATION NUMBER: 60/063540

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063541

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063544

PRIOR FILING DATE: 1997-10-28

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 310

LENGTH: 596

TYPE: PRT

ORGANISM: Homo Sapien

US-10-194-457-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 17

US-10-184-642-310  
/ Sequence 310, Application US/10184642  
/ Publication No. US20030157635A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Baker, Kevin P.  
/ APPLICANT: Chen, Jian  
/ APPLICANT: Desnoyers, Luc  
/ APPLICANT: Goddard, Audrey  
/ APPLICANT: Godowski, Paul J.  
/ APPLICANT: Gurney, Austin L.  
/ APPLICANT: Pan, James  
/ APPLICANT: Smith, Victoria  
/ APPLICANT: Watanabe, Colin K.  
/ APPLICANT: Wood, William I.  
/ APPLICANT: Zhang, Zemin  
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430RIC194

CURRENT FILING DATE: 2002-06-27

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 310

LENGTH: 596

TYPE: PRT

ORGANISM: Homo Sapien

US-10-184-642-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 18

US-10-196-747-310  
/ Sequence 310, Application US/10196747  
/ Publication No. US20030162250A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Baker, Kevin P.  
/ APPLICANT: Chen, Jian  
/ APPLICANT: Desnoyers, Luc  
/ APPLICANT: Goddard, Audrey  
/ APPLICANT: Godowski, Paul J.  
/ APPLICANT: Gurney, Austin L.  
/ APPLICANT: Pan, James  
/ APPLICANT: Smith, Victoria  
/ APPLICANT: Watanabe, Colin K.  
/ APPLICANT: Wood, William I.  
/ APPLICANT: Zhang, Zemin  
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430RIC346

CURRENT FILING DATE: 2002-07-16

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 310

LENGTH: 596

TYPE: PRT

ORGANISM: Homo Sapien

US-10-194-457-310

; ORGANISM: Homo Sapien  
US-10-196-747-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 19

US-10-015-392A-243

; Sequence 243, Application US/10015392A

; Publication No. US20030166901A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; FILE REFERENCE: P2830P1C58

; CURRENT APPLICATION NUMBER: US/10/015,392A

; CURRENT FILING DATE: 2001-12-12

; PRIOR APPLICATION NUMBER: 60/098716

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098723

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098750

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098803

; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/098821

; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/098843

; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/099536

; PRIOR FILING DATE: 1998-09-09

; PRIOR APPLICATION NUMBER: 60/099596

; PRIOR FILING DATE: 1998-09-09

; PRIOR APPLICATION NUMBER: 60/099598

; PRIOR FILING DATE: 1998-09-09

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 477

; SEQ ID NO 243

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-015-392A-243

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 20

US-10-017-253A-243

; Sequence 243, Application US/10017253A

; Publication No. US20030166055A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; FILE REFERENCE: P2830P1C62

; CURRENT APPLICATION NUMBER: US/10/017,253A

; CURRENT FILING DATE: 2001-12-13

; PRIOR APPLICATION NUMBER: 60/098716

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098723

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098749

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098750

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098803

; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/098821

; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/098843

; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/099536

; PRIOR FILING DATE: 1998-09-09

; PRIOR APPLICATION NUMBER: 60/099596

; PRIOR FILING DATE: 1998-09-09

; PRIOR APPLICATION NUMBER: 60/099598

; PRIOR FILING DATE: 1998-09-09

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 477

; SEQ ID NO 243

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-017-253A-243

Query Match 2.0%; Score 8; DB 12; Length 596;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88

Db 39 SSGASTAT 46

## RESULT 21

US-10-173-689-310

; Sequence 310, Application US/10173689

; Publication No. US20030166104A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James

```

; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C10
; CURRENT APPLICATION NUMBER: US/10/173,689
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-689-310

```

```

Query Match          2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

```

```

RESULT 22
US-10-173-690-310
; Sequence 310, Application US/10173690
; Publication No. US20030166105A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C9
; CURRENT APPLICATION NUMBER: US/10/173,690
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-690-310

```

```

Query Match          2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

```

```

RESULT 23
US-10-173-691-310
; Sequence 310, Application US/10173691
; Publication No. US20030166106A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey

```

```

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C13
; CURRENT APPLICATION NUMBER: US/10/173,691
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-691-310

```

```

Query Match          2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

```

```

RESULT 24
US-10-173-692-310
; Sequence 310, Application US/10173692
; Publication No. US20030166188A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C20
; CURRENT APPLICATION NUMBER: US/10/173,692
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-692-310

```

```

Query Match          2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

```

```

RESULT 25
US-10-173-694-310
; Sequence 310, Application US/10173694
; Publication No. US20030166107A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

```

```

; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C19
; CURRENT APPLICATION NUMBER: US/10/173,694
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-694-310

```

```

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

```

## RESULT 26

```

US-10-173-698-310
; Sequence 310, Application US/10173698
; Publication No. US20030166108A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C12
; CURRENT APPLICATION NUMBER: US/10/173,698
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-698-310

```

```

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

```

## RESULT 27

```

US-10-173-699-310
; Sequence 310, Application US/10173699

```

```

; Publication No. US20030166109A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C8
; CURRENT APPLICATION NUMBER: US/10/173,699
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-699-310

```

```

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

```

## RESULT 28

```

US-10-173-707-310
; Sequence 310, Application US/10173707
; Publication No. US20030166110A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C17
; CURRENT APPLICATION NUMBER: US/10/173,707
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-707-310

```

```

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

```



```

RESULT 29
US-10-174-569-310
; Sequence 310, Application US/10174569
; Publication No. US20030166111A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C39
; CURRENT APPLICATION NUMBER: US/10/174,569
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-569-310

Query Match          2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SSGASTAT 88
Db      39 SSGASTAT 46
|||||

RESULT 30
US-10-174-583-310
; Sequence 310, Application US/10174583
; Publication No. US20030166112A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C22
; CURRENT APPLICATION NUMBER: US/10/174,583
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-583-310

Query Match          2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SSGASTAT 88
Db      39 SSGASTAT 46
|||||

RESULT 29
US-10-174-569-310
; Sequence 310, Application US/10174569
; Publication No. US20030166111A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C39
; CURRENT APPLICATION NUMBER: US/10/174,569
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-569-310

Query Match          2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SSGASTAT 88
Db      39 SSGASTAT 46
|||||

RESULT 30
US-10-174-583-310
; Sequence 310, Application US/10174583
; Publication No. US20030166112A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C22
; CURRENT APPLICATION NUMBER: US/10/174,583
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-583-310

Query Match          2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SSGASTAT 88
Db      39 SSGASTAT 46
|||||

```

```

Db      39 SSGASTAT 46

RESULT 31
US-10-174-587-310
; Sequence 310, Application US/10174587
; Publication No. US20030166113A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C30
; CURRENT APPLICATION NUMBER: US/10/174,587
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-587-310

Query Match          2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SSGASTAT 88
Db      39 SSGASTAT 46
|||||

RESULT 32
US-10-174-589-310
; Sequence 310, Application US/10174589
; Publication No. US20030166114A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C33
; CURRENT APPLICATION NUMBER: US/10/174,589
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-589-310

Query Match          2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 81 SSGASTAT 88  
| | | | |  
Db 39 SSGASTAT 46

## RESULT 33

US-10-174-591-310  
; Sequence 310, Application US/10174591  
; Publication No. US20030166115A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C32  
; CURRENT APPLICATION NUMBER: US/10/174,591  
; CURRENT FILING DATE: 2002-06-18  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-174-591-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
| | | | |  
Db 39 SSGASTAT 46

## RESULT 34

US-10-175-736-310  
; Sequence 310, Application US/10175736  
; Publication No. US20030166117A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C43  
; CURRENT APPLICATION NUMBER: US/10/175,736  
; CURRENT FILING DATE: 2002-06-19  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-175-736-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
| | | | |  
Db 39 SSGASTAT 46

## RESULT 35

US-10-175-742-310  
; Sequence 310, Application US/10175742  
; Publication No. US20030166118A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C59  
; CURRENT APPLICATION NUMBER: US/10/175,742  
; CURRENT FILING DATE: 2002-06-19  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-175-742-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
| | | | |  
Db 39 SSGASTAT 46

## RESULT 36

US-10-175-744-310  
; Sequence 310, Application US/10175744  
; Publication No. US20030166119A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C49  
; CURRENT APPLICATION NUMBER: US/10/175,744  
; CURRENT FILING DATE: 2002-06-19  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
US-10-175-744-310

```
; ORGANISM: Homo Sapien
US-10-175-744-310

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SSGASTAT 88
Db      39 SSGASTAT 46

RESULT 37
US-10-175-745-310
; Sequence 310, Application US/10175745
; Publication No. US20030166120A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C55
; CURRENT APPLICATION NUMBER: US/10/175,745
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-745-310

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SSGASTAT 88
Db      39 SSGASTAT 46

RESULT 38
US-10-175-748-310
; Sequence 310, Application US/10175748
; Publication No. US20030166121A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C62
; CURRENT APPLICATION NUMBER: US/10/175,748
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-748-310

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SSGASTAT 88
Db      39 SSGASTAT 46

RESULT 39
US-10-175-751-310
; Sequence 310, Application US/10175751
; Publication No. US20030166122A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C57
; CURRENT APPLICATION NUMBER: US/10/175,751
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-751-310

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SSGASTAT 88
Db      39 SSGASTAT 46

RESULT 40
US-10-175-754-310
; Sequence 310, Application US/10175754
; Publication No. US20030166123A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C58
; CURRENT APPLICATION NUMBER: US/10/175,754
```

; CURRENT FILING DATE: 2002-06-19  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-175-754-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
| | | | |  
Db 39 SSGASTAT 46

## RESULT 41

US-10-176-480-310  
; Sequence 310, Application US/10176480  
; Publication No. US20030166124A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C96  
; CURRENT APPLICATION NUMBER: US/10/176,480  
; CURRENT FILING DATE: 2002-06-21  
; Prior Application removed - See file Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-480-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
| | | | |  
Db 39 SSGASTAT 46

## RESULT 42

US-10-176-489-310  
; Sequence 310, Application US/10176489  
; Publication No. US20030166125A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3430R1C102  
; CURRENT APPLICATION NUMBER: US/10/176,489  
; CURRENT FILING DATE: 2002-06-21  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-489-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
| | | | |  
Db 39 SSGASTAT 46

## RESULT 43

US-10-176-754-310  
; Sequence 310, Application US/10176754  
; Publication No. US20030166126A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C85  
; CURRENT APPLICATION NUMBER: US/10/176,754  
; CURRENT FILING DATE: 2002-06-20  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-754-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
| | | | |  
Db 39 SSGASTAT 46

## RESULT 44

US-10-176-755-310  
; Sequence 310, Application US/10176755  
; Publication No. US20030166127A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.

```

; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C106
; CURRENT APPLICATION NUMBER: US/10/176,755
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-755-310

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

RESULT 45
US-10-176-759-310
; Sequence 310, Application US/10176759
; Publication No. US20030166128A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C80
; CURRENT APPLICATION NUMBER: US/10/176,759
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-759-310

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

RESULT 46
US-10-176-920-310
; Sequence 310, Application US/10176920
; Publication No. US20030166129A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C120
; CURRENT APPLICATION NUMBER: US/10/176,922
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-920-310

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

RESULT 47
US-10-176-922-310
; Sequence 310, Application US/10176922
; Publication No. US20030166130A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C120
; CURRENT APPLICATION NUMBER: US/10/176,922
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-922-310

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

RESULT 48
US-10-176-924-310
; Sequence 310, Application US/10176924
; Publication No. US20030166131A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc

```



```
; Sequence 310, Application US/10179515
; Publication No. US20030166135A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C131
; CURRENT APPLICATION NUMBER: US/10/179,515
; CURRENT FILING DATE: 2002-06-24
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-179-515-310
```

```
Query Match 2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 81 SSGASTAT 88
Db 39 SSGASTAT 46
```

```
RESULT 53
US-10-017-306A-243
; Sequence 243, Application US/10017306A
; Publication No. US20030170718A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC66
; CURRENT APPLICATION NUMBER: US/10/017,306A
; CURRENT FILING DATE: 2002-06-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 243
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-306A-243
```

```
Query Match 2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 81 SSGASTAT 88
```

```
Db 39 SSGASTAT 46
|||||
```

```
RESULT 54
US-10-063-526-100
; Sequence 100, Application US/10063526
; Publication No. US20030171550A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,526
; CURRENT FILING DATE: 2002-05-02
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 100
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-526-100
```

```
Query Match 2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 81 SSGASTAT 88
Db 39 SSGASTAT 46
|||||
```

```
RESULT 55
US-10-173-702-310
; Sequence 310, Application US/10173702
; Publication No. US20030170793A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C15
; CURRENT APPLICATION NUMBER: US/10/173,702
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-702-310
```

```
Query Match 2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 81 SSGASTAT 88  
 |||||  
 Db 39 SSGASTAT 46

## RESULT 56

US-10-173-703-310  
 ; Sequence 310, Application US/10173703  
 ; Publication No. US20030170794A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P3430R1C16  
 ; CURRENT APPLICATION NUMBER: 2002-06-17  
 ; CURRENT FILING DATE: 2002-06-17  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 310  
 ; LENGTH: 596  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapien  
 US-10-173-703-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
 |||||  
 Db 39 SSGASTAT 46

## RESULT 57

US-10-173-704-310  
 ; Sequence 310, Application US/10173704  
 ; Publication No. US20030170795A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P3430R1C2  
 ; CURRENT APPLICATION NUMBER: US/10/173,704  
 ; CURRENT FILING DATE: 2002-06-17  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 310  
 ; LENGTH: 596  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapien  
 US-10-173-704-310

Query Match 2.0%; Score 8; DB 12; Length 596;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
 |||||  
 Db 39 SSGASTAT 46

## RESULT 58

US-10-174-574-310  
 ; Sequence 310, Application US/10174574  
 ; Publication No. US20030170796A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P3430R1C25  
 ; CURRENT APPLICATION NUMBER: US/10/174,574  
 ; CURRENT FILING DATE: 2002-06-18  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 310  
 ; LENGTH: 596  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapien  
 US-10-174-574-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 SSGASTAT 88  
 |||||  
 Db 39 SSGASTAT 46

## RESULT 59

US-10-176-486-310  
 ; Sequence 310, Application US/10176486  
 ; Publication No. US20030170797A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P3430R1C13  
 ; CURRENT APPLICATION NUMBER: US/10/176,486  
 ; CURRENT FILING DATE: 2002-06-21  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 310  
 ; LENGTH: 596  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapien



US-10-176-486-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

RESULT 60

US-10-176-490-310  
; Sequence 310, Application US/10176490  
; Publication No. US20030170798A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C79  
; CURRENT APPLICATION NUMBER: US/10/176,490  
; CURRENT FILING DATE: 2002-06-20  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-490-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

RESULT 61

US-10-176-752-310  
; Sequence 310, Application US/10176752  
; Publication No. US20030170799A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C101  
; CURRENT APPLICATION NUMBER: US/10/176,752  
; CURRENT FILING DATE: 2002-06-21  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310

; LENGTH: 596

; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-752-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

RESULT 62

US-10-176-981-310  
; Sequence 310, Application US/10176981  
; Publication No. US20030170800A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C73  
; CURRENT APPLICATION NUMBER: US/10/176,981  
; CURRENT FILING DATE: 2002-06-20  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-981-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

RESULT 63

US-10-176-983-310  
; Sequence 310, Application US/10176983  
; Publication No. US20030170801A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C97  
; CURRENT APPLICATION NUMBER: US/10/176,983  
; CURRENT FILING DATE: 2002-06-21

; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-983-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 64

US-10-176-988-310  
; Sequence 310, Application US/10176988  
; Publication No. US20030170802A1

; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430R1C87  
; CURRENT FILING DATE: 2002-06-20

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 310

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-176-988-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 65

US-10-179-517-310

; Sequence 310, Application US/10179517

; Publication No. US20030170805A1

; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE OF INVENTION: ACIDS ENCODING THE SAME

; CURRENT FILING DATE: 2002-06-20

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 310

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo Sapien

; FILE REFERENCE: P3430R1C136  
; CURRENT APPLICATION NUMBER: US/10/179,517  
; CURRENT FILING DATE: 2002-06-24  
; Prior Application removed - See file Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-179-517-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 66

US-10-179-521-310

; Sequence 310, Application US/10179521

; Publication No. US20030170806A1

; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430R1C139

; CURRENT APPLICATION NUMBER: US/10/179,521

; CURRENT FILING DATE: 2002-06-24

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 310

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-179-521-310

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 67

US-10-063-586-100

; Sequence 100, Application US/10063586

; Publication No. US20030176684A1

; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.

; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.

; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE OF INVENTION: ACIDS ENCODING THE SAME

; CURRENT FILING DATE: 2002-06-20

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 310

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo Sapien

ORGANISM: Homo Sapien  
US-10-063-586-100

Publication No. US20030180792A1

; APPLICANT: BOVSTEIN, Desnoyers,  
 ; APPLICANT: Desnoyers, Eaton Dan  
 ; APPLICANT: Eaton Dan

RESULT 68  
US-10-012-064A-243  
; Sequence 243, Application US/10012064A

APPLICANT: GUINCY, RUSSELL D.  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Pan, James

; PRIOR FILING DATE: 1998-09-01

;; PRIOR APPLICATION NUMBER: 60/098750  
;; PRIOR FILING DATE: 1998-09-01  
: PRIOR APPLICATION NUMBER: 60/098803

;; PRIOR FILING DATE: 1998-09-02  
;; PRIOR APPLICATION NUMBER: 60/098821

; PRIOR APPLICATION NUMBER: 60/099815  
 ; PRIOR FILING DATE: 1998-09-10  
 ; PRIOR APPLICATION NUMBER: 60/099816

; PRIOR FILING DATE: 1998-09-10  
 ; PRIOR APPLICATION NUMBER: 60/0999816

PRIOR FILING DATE: 1998-09-10

;  
; PRIOR APPLICATION NUMBER: 60/100385  
; PRIOR FILING DATE: 1998-09-15  
; PRIOR APPLICATION NUMBER: 60/100388  
; PRIOR FILING DATE: 1998-09-15  
; PRIOR APPLICATION NUMBER: 60/100390  
; PRIOR FILING DATE: 1998-09-15  
; PRIOR APPLICATION NUMBER: 60/100584  
; PRIOR FILING DATE: 1998-09-16  
; PRIOR APPLICATION NUMBER: 60/100627  
; PRIOR FILING DATE: 1998-09-16  
; PRIOR APPLICATION NUMBER: 60/100661  
; PRIOR FILING DATE: 1998-09-16  
; PRIOR APPLICATION NUMBER: 60/100662  
; PRIOR FILING DATE: 1998-09-16  
; PRIOR APPLICATION NUMBER: 60/100664  
; PRIOR FILING DATE: 1998-09-16  
; PRIOR APPLICATION NUMBER: 60/100683  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/100684  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/100710  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/100711  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/100848  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: 60/100849  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: 60/100919  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/100930  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/101014  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: 60/101068  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: 60/101071  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: 60/101279  
; PRIOR FILING DATE: 1998-09-22  
; PRIOR APPLICATION NUMBER: 60/101471  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101472  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101474  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101475  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101476  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101477  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101479  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101738  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101741  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101743  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101915  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101916  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/102207  
; PRIOR FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: 60/102240  
; PRIOR FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: 60/102307  
; PRIOR FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: 60/102330  
; PRIOR FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: 60/102331

;  
; PRIOR FILING DATE: 1998-09-29  
; PRIOR APPLICATION NUMBER: 60/102484  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 60/102487  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 60/102570  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 60/102571  
; PRIOR FILING DATE: 1998-09-30  
; PRIOR APPLICATION NUMBER: 60/102684  
; PRIOR FILING DATE: 1998-10-01  
; PRIOR APPLICATION NUMBER: 60/102687  
; PRIOR FILING DATE: 1998-10-01  
; PRIOR APPLICATION NUMBER: 60/102965  
; PRIOR FILING DATE: 1998-10-02  
; PRIOR APPLICATION NUMBER: 60/103258  
; PRIOR FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: 60/103314  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103315  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103328  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103395  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103396  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103401  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 60/103449  
; PRIOR FILING DATE: 1998-10-06  
; PRIOR APPLICATION NUMBER: 60/103633  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/103678  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/103679  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/103711  
; PRIOR FILING DATE: 1998-10-08  
; PRIOR APPLICATION NUMBER: 60/104257  
; PRIOR FILING DATE: 1998-10-14  
; PRIOR APPLICATION NUMBER: 60/104987  
; PRIOR FILING DATE: 1998-10-20  
; PRIOR APPLICATION NUMBER: 60/105000  
; PRIOR FILING DATE: 1998-10-20  
; PRIOR APPLICATION NUMBER: 60/105002  
; PRIOR FILING DATE: 1998-10-20  
; PRIOR APPLICATION NUMBER: 60/105104  
; PRIOR FILING DATE: 1998-10-21  
; PRIOR APPLICATION NUMBER: 60/105169  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: 60/105266  
; PRIOR FILING DATE: 1998-10-22  
; PRIOR APPLICATION NUMBER: 60/105693  
; PRIOR FILING DATE: 1998-10-26  
; PRIOR APPLICATION NUMBER: 60/105694  
; PRIOR FILING DATE: 1998-10-26  
; PRIOR APPLICATION NUMBER: 60/105807  
; PRIOR FILING DATE: 1998-10-27  
; PRIOR APPLICATION NUMBER: 60/105881  
; PRIOR FILING DATE: 1998-10-27  
; PRIOR APPLICATION NUMBER: 60/105882  
; PRIOR FILING DATE: 1998-10-27  
; PRIOR APPLICATION NUMBER: 60/106023  
; PRIOR FILING DATE: 1998-10-28

Query Match 2.0%; Score 8; DB 12; Length 596;

Best Local Similarity 100.0%; Pred. No. 1.2e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0;

QY 81 SSGASTAT 88  
Db 39 SSGASTAT 46

## RESULT 70

US-10-063-510-100  
; Sequence 100, Application US/10063510  
; Publication No. US20030180837A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063,510  
; Prior Filing Date: 2002-05-01  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 100  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-510-100

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
DB 39 SSGASTAT 46

## RESULT 71

US-10-063-514-100  
; Sequence 100, Application US/10063514  
; Publication No. US20030181707A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063,514  
; Prior Filing Date: 2002-05-01  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 100  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-514-100

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
DB 39 SSGASTAT 46

## RESULT 72

US-10-063-516-100  
; Sequence 100, Application US/10063516  
; Publication No. US20030181708A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063,516  
; Prior Filing Date: 2002-05-01  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 100  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-516-100

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
DB 39 SSGASTAT 46

## RESULT 73

US-10-063-523-100  
; Sequence 100, Application US/10063523  
; Publication No. US20030181636A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063,523  
; Prior Filing Date: 2002-05-02  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 100  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-523-100

Query Match 2.0%; Score 8; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SSGASTAT 88  
DB 39 SSGASTAT 46

Job time : 36.0556 secs

```

RESULT 74
US-10-063-527-100
; Sequence 100, Application US/10063527
; Publication No. US20030181637A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,527
; CURRENT FILING DATE: 2002-05-02
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 100
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-527-100

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

```

```

RESULT 75
US-10-063-528-100
; Sequence 100, Application US/10063528
; Publication No. US20030181666A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,528
; CURRENT FILING DATE: 2002-05-02
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 100
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-528-100

Query Match      2.0%; Score 8; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      81 SSGASTAT 88
Db      39 SSGASTAT 46

```

Search completed: December 17, 2003, 06:42:59

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:32:02 ; Search time 14.1462 Seconds  
(without alignments)  
1211.344 Million cell updates/sec

Title: US-09-805-550-2

Perfect score: 405

Sequence: 1 MKNVTKLGTNFEIASPD.....ELTANYLLDHGHEFDQQQ 405

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5A COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PTUS COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	405	100.0	405	3	US-09-413-574-2
2	13	3.2	368	3	US-09-413-574-4
3	8	2.0	23	2	US-08-492-027A-3
4	8	2.0	117	6	5242821-12
5	8	2.0	177	4	US-09-252-991A-28452
6	8	2.0	180	4	US-09-252-991A-18890
7	8	2.0	212	4	US-09-252-991A-24512
8	8	2.0	272	2	US-08-492-027A-4
9	8	2.0	495	4	US-09-252-991A-31949
10	8	2.0	568	4	US-09-252-991A-23264
11	8	2.0	720	4	US-09-252-991A-31915
12	8	2.0	892	4	US-09-585-858-16
13	8	2.0	898	2	US-08-808-982-5
14	8	2.0	898	3	US-09-306-902A-5
15	7	1.7	62	1	US-08-428-091-1
16	7	1.7	78	1	US-07-881-075-5
17	7	1.7	78	1	US-08-120-827-5
18	7	1.7	78	1	US-08-478-675-5
19	7	1.7	85	1	US-08-428-091-4
20	7	1.7	104	4	US-09-134-001C-4716
21	7	1.7	150	4	US-09-252-991A-30419
22	7	1.7	155	4	US-09-252-991A-20273
23	7	1.7	159	4	US-09-252-991A-18088
24	7	1.7	179	4	US-09-252-991A-26223
25	7	1.7	191	3	US-08-974-022-52
26	7	1.7	191	3	US-08-795-445A-52
27	7	1.7	191	3	US-08-795-447A-52

28	7	1.7	191	3	US-08-974-186-52	Sequence 52, Appl
29	7	1.7	191	3	US-08-795-446B-52	Sequence 52, Appl
30	7	1.7	191	4	US-08-706-945D-139	Sequence 139, App
31	7	1.7	221	4	US-09-904-615-155	Sequence 155, App
32	7	1.7	256	1	US-08-236-918A-6	Sequence 6, Appl
33	7	1.7	256	4	US-09-150-864A-6	Sequence 6, Appl
34	7	1.7	256	4	US-08-012-269A-2	Sequence 2, Appl
35	7	1.7	256	5	PCT-US96-03965-2	Sequence 2, Appl
36	7	1.7	258	4	US-09-252-991A-37264	Sequence 27264, A
37	7	1.7	272	4	US-09-252-991A-17461	Sequence 17461, A
38	7	1.7	281	1	US-08-397-633A-75	Sequence 75, Appl
39	7	1.7	281	4	US-09-252-991A-26761	Sequence 26761, A
40	7	1.7	287	1	US-08-397-633A-76	Sequence 76, Appl
41	7	1.7	296	4	US-09-252-991A-22293	Sequence 22293, A
42	7	1.7	303	4	US-09-495-406-17	Sequence 17, Appl
43	7	1.7	341	4	US-09-134-001C-3568	Sequence 3568, Ap
44	7	1.7	343	4	US-09-252-991A-24458	Sequence 24458, A
45	7	1.7	360	4	US-09-107-532A-6666	Sequence 6666, Ap
46	7	1.7	364	4	US-09-252-991A-26295	Sequence 26295, A
47	7	1.7	382	4	US-09-252-991A-23727	Sequence 23727, A
48	7	1.7	391	3	US-09-377-557-18	Sequence 18, Appl
49	7	1.7	406	4	US-09-252-991A-21075	Sequence 21075, A
50	7	1.7	415	4	US-09-252-991A-27669	Sequence 27669, A
51	7	1.7	420	4	US-09-252-991A-22253	Sequence 22253, A
52	7	1.7	423	4	US-09-252-991A-23618	Sequence 23618, A
53	7	1.7	448	3	US-08-476-509B-2	Sequence 2, Appl
54	7	1.7	465	3	US-08-948-997-6	Sequence 6, Appl
55	7	1.7	465	3	US-09-348-817A-6	Sequence 6, Appl
56	7	1.7	465	4	US-09-722-292-6	Sequence 6, Appl
57	7	1.7	479	4	US-09-252-991A-23144	Sequence 23144, A
58	7	1.7	486	3	US-08-348-518C-2	Sequence 2, Appl
59	7	1.7	495	4	US-09-252-991A-37501	Sequence 37501, A
60	7	1.7	519	4	US-09-252-991A-16737	Sequence 16737, A
61	7	1.7	534	4	US-09-252-991A-22537	Sequence 22537, A
62	7	1.7	549	4	US-09-252-991A-26327	Sequence 26327, A
63	7	1.7	561	4	US-09-252-991A-23080	Sequence 23080, A
64	7	1.7	567	4	US-09-205-258-573	Sequence 573, App
65	7	1.7	591	3	US-09-199-290-7	Sequence 7, Appl
66	7	1.7	617	4	US-09-252-991A-23472	Sequence 23472, A
67	7	1.7	618	3	US-09-198-290-34	Sequence 34, Appl
68	7	1.7	619	4	US-09-252-991A-21585	Sequence 21585, A
69	7	1.7	657	4	US-09-252-991A-28001	Sequence 28001, A
70	7	1.7	661	2	US-08-795-868-14	Sequence 14, Appl
71	7	1.7	661	4	US-09-303-069-14	Sequence 14, Appl
72	7	1.7	661	4	US-09-134-250-14	Sequence 14, Appl
73	7	1.7	749	4	US-09-252-991A-20752	Sequence 20752, A
74	7	1.7	798	4	US-09-252-991A-23774	Sequence 23774, A
75	7	1.7	802	4	US-09-252-991A-25050	Sequence 25050, A
76	7	1.7	803	3	US-09-118-442-2	Sequence 2, Appl
77	7	1.7	803	3	US-09-677-064-2	Sequence 2, Appl
78	7	1.7	848	4	US-09-252-991A-25143	Sequence 25143, A
79	7	1.7	872	1	US-08-491-357-3	Sequence 3, Appl
80	7	1.7	872	3	US-08-968-633-3	Sequence 3, Appl
81	7	1.7	872	3	US-09-196-466-3	Sequence 3, Appl
82	7	1.7	872	5	PCT-US96-10823-3	Sequence 3, Appl
83	7	1.7	877	1	US-08-397-633A-54	Sequence 54, Appl
84	7	1.7	906	4	US-09-252-991A-31458	Sequence 31458, A
85	7	1.7	952	4	US-09-252-991A-32183	Sequence 32183, A
86	7	1.7	1004	3	US-08-916-352-2	Sequence 2, Appl
87	7	1.7	1011	4	US-09-252-991A-32419	Sequence 32419, A
88	7	1.7	1042	4	US-09-512-250C-32	Sequence 32, Appl
89	7	1.7	1049	2	US-08-817-090B-2	Sequence 2, Appl
90	7	1.7	1050	2	US-08-817-090B-4	Sequence 4, Appl
91	7	1.7	1096	4	US-09-252-991A-19328	Sequence 19328, A
92	7	1.7	1101	2	US-08-916-917-14	Sequence 14, Appl
93	7	1.7	1101	3	US-09-225-170-14	Sequence 14, Appl
94	7	1.7	1102	2	US-08-916-917-4	Sequence 4, Appl
95	7	1.7	1102	2	US-08-972-631-4	Sequence 4, Appl
96	7	1.7	1102	2	US-08-972-629-4	Sequence 4, Appl
97	7	1.7	1102	2	US-08-972-630-4	Sequence 4, Appl
98	7	1.7	1102	2	US-08-672-211-4	Sequence 4, Appl
99	7	1.7	1102	3	US-09-225-170-4	Sequence 4, Appl
100	7	1.7	1246	4	US-09-252-991A-23140	Sequence 23140, A

## ALIGNMENTS

## RESULT 1

US-09-413-574-2  
; Sequence 2, Application US/09413574  
; Patent No. 6235972  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; APPLICANT: Taghiani, Laura  
; TITLE OF INVENTION: Maize Rad23 Genes and Uses Thereof  
; FILE REFERENCE: 0964  
; CURRENT APPLICATION NUMBER: US/09/413,574  
; CURRENT FILING DATE: 1999-10-06  
; EARLIER APPLICATION NUMBER: 60/109,728  
; EARLIER FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 405  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-413-574-2

Query Match 100.0%; Score 405; DB 3; Length 405;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKLVKTLKGNFRIEASPDASVADVKRIETTTQGSTYRADQQMLIYQGKILKDETTLE	60
Db	1	MKLVKTLKGNFRIEASPDASVADVKRIETTTQGSTYRADQQMLIYQGKILKDETTLE	60
Qy	61	SNGVAENFLVIMLSKAKASSGASTATTAKAPATLAQPAAPVAPAAASVARTPTQAPVAT	120
Db	61	SNGVAENFLVIMLSKAKASSGASTATTAKAPATLAQPAAPVAPAAASVARTPTQAPVAT	120
Qy	121	AETAPPSVQQAAPAAATVAAATDDADVTYSOASNLVFGNNLEBTIQQILDMGGGTWERDTV	180
Db	121	AETAPPSVQQAAPAAATVAAATDDADVTYSOASNLVFGNNLEBTIQQILDMGGGTWERDTV	180
Qy	181	VRALRAAYNNPERAIDLYSGIPENVEAQPVARAPAAAGQQTNOQAASPAQPAVALPVQPS	240
Db	181	VRALRAAYNNPERAIDLYSGIPENVEAQPVARAPAAAGQQTNOQAASPAQPAVALPVQPS	240
Qy	241	PASAGPNANPLNLPFGVPSGSGNPGVVPAGSGALDALROLPOFQALLQVLQANPQILQ	300
Db	241	PASAGPNANPLNLPFGVPSGSGNPGVVPAGSGALDALROLPOFQALLQVLQANPQILQ	300
Qy	301	PMQLGKQNPQILRLIQENQAEFLRLVNESPEGGPGGNILGQLAAAVPOTLTVTTPERE	360
Db	301	PMQLGKQNPQILRLIQENQAEFLRLVNESPEGGPGGNILGQLAAAVPOTLTVTTPERE	360
Qy	361	AIORLEGMGFNRELVLVFFACNKKDEELTANYLLDHGHEFDQDQ 405	
Db	361	AIORLEGMGFNRELVLVFFACNKKDEELTANYLLDHGHEFDQDQ 405	

## RESULT 2

US-09-413-574-4  
; Sequence 4, Application US/09413574  
; Patent No. 6235972  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; APPLICANT: Taghiani, Laura  
; TITLE OF INVENTION: Maize Rad23 Genes and Uses Thereof  
; FILE REFERENCE: 0964  
; CURRENT APPLICATION NUMBER: US/09/413,574  
; CURRENT FILING DATE: 1999-10-06  
; EARLIER APPLICATION NUMBER: 60/109,728  
; EARLIER FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-413-574-4

Query Match 3.2%; Score 13; DB 3; Length 368;  
Best Local Similarity 100.0%; Pred. No. 0.00061;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	182	RALRAAYNNPERA	194
Db	167	RALRAAYNNPERA	179

## RESULT 3

US-08-492-027A-3  
; Sequence 3, Application US/08492027A  
; Patent No. 5912333  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shoichi  
; APPLICANT: Burnell, James N  
; TITLE OF INVENTION: DNA Encoding Carbonic Anhydrase  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch and Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/492,027A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr, Gerald M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 0760-206  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-492-027A-3

Query Match 2.0%; Score 8; DB 2; Length 23;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	100	AAPVAPAA	107
Db	1	AAPVAPAA	8

## RESULT 4

5242821-12  
; Patent No. 5242821  
; APPLICANT: PALVA, LIKKA;SIBAKOV, MERVI  
; TITLE OF INVENTION: LACTOCOCCUS PROMOTER AND SIGNAL  
; SEQUENCES FOR EXPRESSION IN BACTERIA  
; NUMBER OF SEQUENCES: 27



; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/377,450  
; FILING DATE: 10-JUL-1989  
; SEQ ID NO:12:  
; LENGTH: 117  
5242821-12

Query Match 2.0%; Score 8; DB 6; Length 117;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 SKAKASSS 82  
Db 72 SKAKASSS 79

## RESULT 5

US-09-252-991A-28452  
; Sequence 28452, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28452  
; LENGTH: 177  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28452

Query Match 2.0%; Score 8; DB 4; Length 177;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 APVAPAS 108  
Db 36 APVAPAS 43

## RESULT 6

US-09-252-991A-18890  
; Sequence 18890, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18890  
; LENGTH: 180  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18890

Query Match 2.0%; Score 8; DB 4; Length 180;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 APAATVAA 140  
Db 115 APAATVAA 122

## RESULT 7

US-09-252-991A-24512  
; Sequence 24512, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24512  
; LENGTH: 212  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (43)  
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.  
US-09-252-991A-24512

Query Match 2.0%; Score 8; DB 4; Length 212;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 PAAPVAPA 106  
Db 140 PAAPVAPA 147

## RESULT 8

US-08-492-027A-4  
; Sequence 4, Application US/08492027A  
; Patent No. 5912333  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shoichi  
; TITLE OF INVENTION: DNA Encoding Carbonic Anhydrase  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch and Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/492,027A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr, Gerald M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 0760-206  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 272 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-492-027A-4

Query Match 2.0%; Score 8; DB 2; Length 272;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 AAPVAPAA 107  
Db 63 AAPVAPAA 70

## RESULT 9

US-09-252-991A-31949  
Sequence 31949, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 31949  
LENGTH: 495  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31949

Query Match 2.0%; Score 8; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 ARAPAAQ 219  
Db 131 ARAPAAQ 138

## RESULT 10

US-09-252-991A-23264  
Sequence 23264, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 23264  
LENGTH: 568  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-23264

Query Match 2.0%; Score 8; DB 4; Length 568;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AQAAPVA 104  
Db 257 AQAAPVA 264

## RESULT 11

US-09-252-991A-31915  
Sequence 31915, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 31915  
LENGTH: 720  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31915

Query Match 2.0%; Score 8; DB 4; Length 720;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 AAPVAPAA 107  
Db 465 AAPVAPAA 472

## RESULT 12

US-09-585-858-16  
Sequence 16, Application US/09585858  
Patent No. 6492161

GENERAL INFORMATION:

APPLICANT: Sigridur Hjorleifsdottir  
APPLICANT: Gudmundur O. Hreggvidsson  
APPLICANT: Olafur H. Fridjonsson  
APPLICANT: Arnthor Aevarsson  
APPLICANT: Jakob K. Kristjansson  
TITLE OF INVENTION: Bacteriophage RM378 of a Thermophilic  
FILE REFERENCE: 2739.1001-001  
CURRENT APPLICATION NUMBER: US/09/585,858  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: 60/137,120  
PRIOR FILING DATE: 1999-06-02  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16  
LENGTH: 892  
TYPE: PRT  
ORGANISM: Herpes simplex virus (type 1/strain 17)

US-09-585-858-16

Query Match 2.0%; Score 8; DB 4; Length 892;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 TLAQPAAP 102  
Db 38 TLAQPAAP 45

## RESULT 13

US-08-808-982-5  
Sequence 5, Application US/08808982

; Patent No. 5939271  
; GENERAL INFORMATION:  
; APPLICANT: Tessier-Lavigne, Marc  
; APPLICANT: Leonardo, E. David  
; APPLICANT: Hink, Lindsey  
; APPLICANT: Masu, Masaayuki  
; APPLICANT: Kazuko, Keino-Masu  
; TITLE OF INVENTION: Netrin Receptors  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/808,982  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UC96-217  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 898 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; US-08-808-982-5

Query Match 2.0%; Score 8; DB 2; Length 898;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 LGQLAAAV 348  
| | | | |  
Db 873 LGQLAAAV 880

RESULT 14  
US-09-306-902A-5  
; Sequence 5, Application US/09306902A  
; Patent No. 627585  
; GENERAL INFORMATION:  
; APPLICANT: Tessier-Lavigne, Marc  
; APPLICANT: Leonardo, E. David  
; APPLICANT: Hink, Lindsey  
; APPLICANT: Masu, Masaayuki  
; APPLICANT: Kazuko, Keino-Masu  
; TITLE OF INVENTION: Netrin Receptors  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/428,091  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP PCT/EP94/02805  
; FILING DATE: 24-AUG-1994  
; APPLICATION NUMBER: CH 2628/93-7  
; FILING DATE: 03-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FANUCCI, ALLAN A  
; REGISTRATION NUMBER: 30256  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790 9090  
; TELEFAX: 212 869 8864  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 62 amino acids  
; TYPE: amino acid

; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/306,902A  
; FILING DATE: 07-May-1999  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UC96-217  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 898 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; US-09-306-902A-5

Query Match 2.0%; Score 8; DB 3; Length 898;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 LGQLAAAV 348  
| | | | |  
Db 873 LGQLAAAV 880

RESULT 15  
US-08-428-091-1  
; Sequence 1, Application US/08428091  
; Patent No. 568380  
; GENERAL INFORMATION:  
; APPLICANT: GERMOND, JACQUES.-EDOUARD  
; APPLICANT: MARCISSET, OLIVIER  
; APPLICANT: MOLLET, BEAT  
; TITLE OF INVENTION: BACTERIOICINS OF STREPTOCOCCUS  
; TITLE OF INVENTION: THERMOPHILUS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 AVENUE OF THE AMERICA  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/428,091  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP PCT/EP94/02805  
; FILING DATE: 24-AUG-1994  
; APPLICATION NUMBER: CH 2628/93-7  
; FILING DATE: 03-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FANUCCI, ALLAN A  
; REGISTRATION NUMBER: 30256  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790 9090  
; TELEFAX: 212 869 8864  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 62 amino acids  
; TYPE: amino acid

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus thermophilus
; STRAIN: CNM I-1351
US-08-428-091-1

Query Match      1.7%  Score 7; DB 1; Length 62;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      270 GAGSGAL 276
Db      17 GAGSGAL 23

RESULT 16
US-07-881-075-5
; Sequence 5, Application US/07881075
; Patent No. 5444149
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 19920511
; APPLICATION NUMBER: US/07/881.075
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5444149man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-154-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-07-881-075-5

Query Match      1.7%  Score 7; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      284 QFQALLQ 290
Db      37 QFQALLQ 43

RESULT 17
US-08-120-827-5
; Sequence 5, Application US/08120827
```

```
; Patent No. 5525495
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 15-SEP-1993
; APPLICATION NUMBER: US/08/120,827
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5525495man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)413-3000
; TELEFAX: (703)413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-120-827-5

Query Match      1.7%  Score 7; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      284 QFQALLQ 290
Db      37 QFQALLQ 43

RESULT 18
US-08-478-675-5
; Sequence 5, Application US/08478675
; Patent No. 5773246
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,675  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/120,827  
FILING DATE: 15-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Orlon, No. 577346man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 714-158-0 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)413-3000  
TELEFAX: (703)413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-478-675-5

Query Match 1.7%; Score 7; DB 1; Length 78;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 QFOALLQ 290  
DB 37 QFOALLQ 43

RESULT 19  
US-08-428-091-4  
Sequence 4, Application US/08428091  
Patent No. 5683890  
GENERAL INFORMATION:  
APPLICANT: GERMOND, JACQUES.-EDOUARD  
APPLICANT: MARCISSET, OLIVIER  
APPLICANT: MOLLET, BEAT  
TITLE OF INVENTION: BACTERIOCINS OF STREPTOCOCCUS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 AVENUE OF THE AMERICA  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/428,091  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP PCT/EP94/02805  
FILING DATE: 24-AUG-1994  
APPLICATION NUMBER: CH 2628/93-7  
FILING DATE: 03-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: FANUCCI, ALLAN A  
REGISTRATION NUMBER: 30256  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790 9090  
TELEFAX: 212 869 8864

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 85 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-428-091-4

Query Match 1.7%; Score 7; DB 1; Length 85;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 GAGSGAL 276  
DB 40 GAGSGAL 46

RESULT 20  
US-09-134-001C-4716  
Sequence 4716, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4716  
LENGTH: 104  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4716

Query Match 1.7%; Score 7; DB 4; Length 104;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 SGASTAT 88  
DB 70 SGASTAT 76

RESULT 21  
US-09-252-991A-30419  
Sequence 30419, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 30419  
LENGTH: 150  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30419

Query Match 1.7%; Score 7; DB 4; Length 150;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 AQAAPV 103  
|||||  
Db 25 AQAAPV 31

## RESULT 22

US-09-252-991A-20273  
; Sequence 20273, Application US/09252991A  
; Patent No. 6551795

## ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20273  
; LENGTH: 155  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20273

Query Match 1.7%; Score 7; DB 4; Length 155;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 ASTATTA 90  
|||||  
Db 3 ASTATTA 9

## RESULT 23

US-09-252-991A-18088  
; Sequence 18088, Application US/09252991A  
; Patent No. 6551795

## ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18088  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18088

Query Match 1.7%; Score 7; DB 4; Length 159;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 TAETAPP 126  
|||||  
Db 89 TAETAPP 95

## RESULT 24

US-09-252-991A-26223  
; Sequence 26223, Application US/09252991A  
; Patent No. 6551795

## ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26223  
; LENGTH: 179  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26223

Query Match 1.7%; Score 7; DB 4; Length 179;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 PAATVAA 140  
|||||  
Db 95 PAATVAA 101

## RESULT 25

US-08-974-022-52  
; Sequence 52, Application US/08974022  
; Patent No. 6015938

## ; GENERAL INFORMATION:

; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Denavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/974,022  
; FILING DATE: 12-DEC-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/577,788  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 191 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-974-022-52

Query Match 1.7%; Score 7; DB 3; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 332 PEGGPG 338

Db 175 PEGGPGG 181  
|||||

## RESULT 26

US-08-795-445A-52  
; Sequence 52, Application US/08795445A  
; Patent No. 6284485

## GENERAL INFORMATION:

APPLICANT: Boyle, William J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: OSTEOPROTEGERIN  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,445A  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/577,788  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-795-445A-52

Query Match 1.7%; Score 7; DB 3; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 PEGGPGG 338  
|||||

## RESULT 27

US-08-795-447A-52  
; Sequence 52, Application US/08795447A  
; Patent No. 6284728

## GENERAL INFORMATION:

APPLICANT: Boyle, William J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: Osteoprotegerin  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: One Amgen Center Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91362-1789

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,447A  
FILING DATE:

## CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378D2  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-795-447A-52

Query Match 1.7%; Score 7; DB 3; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 PEGGPGG 338  
|||||

## RESULT 28

US-08-974-186-52  
; Sequence 52, Application US/08974186  
; Patent No. 6284740

GENERAL INFORMATION:  
APPLICANT: Boyle, William J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: OSTEOPROTEGERIN  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,186  
FILING DATE:

## CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/577,788  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-974-186-52

Query Match 1.7%; Score 7; DB 3; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.2e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;

QY 332 PEGGPGG 338  
Db 175 PEGGPGG 181

## RESULT 29

US-08-795-446B-52  
; Sequence 52, Application US/08795446B  
; Patent No. 6288032  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/795,446B  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/577,788  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 191 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-795-446B-52

Query Match 1.7%; Score 7; DB 3; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 PEGGPGG 338  
Db 175 PEGGPGG 181

## RESULT 30

US-08-706-945D-139  
; Sequence 139, Application US/08706945D  
; Patent No. 6369027  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William  
; APPLICANT: Lacey, David  
; APPLICANT: Calzone, Frank  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: Osteoprotegerin  
; FILE REFERENCE: A-378CIP  
; CURRENT APPLICATION NUMBER: US/08/706,945D  
; CURRENT FILING DATE: 1996-09-03  
; PRIOR APPLICATION NUMBER: 08/577,788

; PRIOR FILING DATE: 1995-12-22  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 139  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-706-945D-139

Query Match 1.7%; Score 7; DB 4; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 PEGGPGG 338  
Db 175 PEGGPGG 181

## RESULT 31

US-09-904-615-155  
; Sequence 155, Application US/09904615  
; Patent No. 6566325  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 49 Human Secreted Proteins  
; FILE REFERENCE: P2032P1  
; CURRENT APPLICATION NUMBER: US/09/904,615  
; CURRENT FILING DATE: 2001-07-16  
; PRIOR APPLICATION NUMBER: 09/511,554  
; PRIOR FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 60/097,917  
; PRIOR FILING DATE: 1998-08-25  
; PRIOR APPLICATION NUMBER: 60/098,634  
; PRIOR FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 170  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 155  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-904-615-155

Query Match 1.7%; Score 7; DB 4; Length 221;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 PGAGSGA 275  
Db 35 PGAGSGA 41

## RESULT 32

US-08-236-918A-6  
; Sequence 6, Application US/08236918A  
; Patent No. 5674704  
; GENERAL INFORMATION:  
; APPLICANT: Alderson, Mark R.  
; APPLICANT: Goodwin, Raymond G.  
; APPLICANT: Smith, Craig A.  
; TITLE OF INVENTION: Cytokine Designated 4-1BB Ligand  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kathryn A. Anderson, Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: US  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Power Macintosh  
; OPERATING SYSTEM: Apple 7.5.3  
; SOFTWARE: Microsoft Word, Version #6.0.1

Query Match 1.7%; Score 7; DB 3; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 PEGGPGG 338  
Db 175 PEGGPGG 181



;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/236,918A  
;/ FILING DATE: 06-May-1994  
;/ CLASSIFICATION: 435  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: US 08/060,843  
;/ FILING DATE: 07-May-1993  
;/ CLASSIFICATION: 435  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Anderson, Kathryn A.  
;/ REGISTRATION NUMBER: 32,172  
;/ REFERENCE/DOCKET NUMBER: 2801-B  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (206) 587-0430  
;/ TELEFAX: (206) 233-0644  
;/ INFORMATION FOR SEQ ID NO: 6:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 256 amino acids  
;/ TYPE: amino acid  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: protein  
US-08-236-918A-6

Query Match 1.7%; Score 7; DB 1; Length 256;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 PEGGPGG 338  
DB 175 PEGGPGG 181

RESULT 33  
US-09-150-864A-6  
;/ Sequence 6, Application US/09150864A  
;/ Patent No. 6355779  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Alderson, Mark R.  
;/ APPLICANT: Goodwin, Raymond G.  
;/ APPLICANT: Smith, Craig A.  
;/ TITLE OF INVENTION: Cytokine Designated 4-1BB Ligand and Human Receptor  
;/ FILE REFERENCE: 2801-B  
;/ CURRENT APPLICATION NUMBER: US/09/150,864A  
;/ CURRENT FILING DATE: 1998-09-10  
;/ PRIOR APPLICATION NUMBER: 08/060,843  
;/ PRIOR FILING DATE: 1993-05-07  
;/ NUMBER OF SEQ ID NOS: 18  
;/ SOFTWARE: PatentIn Ver. 2.0  
;/ SEQ ID NO 6  
;/ LENGTH: 256  
;/ TYPE: PRT  
;/ ORGANISM: Mus sp. (clone: mu4-1BB)  
US-09-150-864A-6

Query Match 1.7%; Score 7; DB 4; Length 256;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 PEGGPGG 338  
DB 175 PEGGPGG 181

RESULT 34  
US-08-012-269A-2  
;/ Sequence 2, Application US/08012269A  
;/ Patent No. 6362325  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Kwon, Byoung S.  
;/ TITLE OF INVENTION: MURINE 4-1BB GENE  
;/ FILE REFERENCE: 740.009U1  
;/ CURRENT APPLICATION NUMBER: US/08/012,269A

;/ CURRENT FILING DATE: 1993-02-01  
;/ PRIOR APPLICATION NUMBER: US 07/922,996  
;/ PRIOR FILING DATE: 1992-07-30  
;/ PRIOR APPLICATION NUMBER: US 07/267,572  
;/ PRIOR FILING DATE: 1988-11-07  
;/ NUMBER OF SEQ ID NOS: 13  
;/ SOFTWARE: FastSeq for Windows Version 4.0  
;/ SEQ ID NO 2  
;/ LENGTH: 256  
;/ TYPE: PRT  
;/ ORGANISM: Mus musculus  
US-08-012-269A-2

Query Match 1.7%; Score 7; DB 4; Length 256;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 PEGGPGG 338  
DB 175 PEGGPGG 181

RESULT 35  
PCT-US96-03965-2  
;/ Sequence 2, Application PC/TUS9603965  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Kwon, Byoung Se  
;/ APPLICANT: Kang, Chang-Yuil  
;/ TITLE OF INVENTION: Monoclonal antibody against human  
;/ TITLE OF INVENTION: receptor 4-1BB  
;/ NUMBER OF SEQUENCES: 10  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: Barnard, Brown & Michaels  
;/ STREET: 306 East State Street, Suite 220  
;/ CITY: Ithaca  
;/ STATE: NY  
;/ COUNTRY: USA  
;/ ZIP: 14850  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: PCT/US96/03965  
;/ FILING DATE:  
;/ CLASSIFICATION:  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: US 08/122,796  
;/ FILING DATE: 16-SEP-1993  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: US 08/012,269  
;/ FILING DATE: 01-FEB-1993  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: US 07/922,996  
;/ FILING DATE: 30-JUL-1992  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: US 07/267,577  
;/ FILING DATE: 07-NOV-1988  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Michaels, Christopher A  
;/ REGISTRATION NUMBER: 34,390  
;/ REFERENCE/DOCKET NUMBER: KW05  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: 607-273-1711  
;/ TELEFAX: 607-273-2609  
;/ INFORMATION FOR SEQ ID NO: 2:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 256 amino acids  
;/ TYPE: amino acid  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: protein  
PCT-US96-03965-2

Query Match 1.7%; Score 7; DB 5; Length 256;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 332 PEGPGG 338  
Db 175 PEGPGG 181  
|||||

RESULT 36  
US-09-252-991A-27264  
; Sequence 27264, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27264  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27264

Query Match 1.7%; Score 7; DB 4; Length 258;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 APAATVA 139  
Db 191 APAATVA 197  
|||||

RESULT 37  
US-09-252-991A-17461  
; Sequence 17461, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17461  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17461

Query Match 1.7%; Score 7; DB 4; Length 272;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 226 ASPAQPA 232  
Db 21 ASPAQPA 27  
|||||

RESULT 38

US-08-397-633A-75  
; Sequence 75, Application US/08397633A  
; Patent No. 5773577  
; GENERAL INFORMATION:  
; APPLICANT: Cappello, Joseph  
; TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE  
; TITLE OF INVENTION: OF ENZYMIC CROSS-LINKING  
; NUMBER OF SEQUENCES: 105  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/397,633A  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Bertram I  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 75:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 281 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-397-633A-75

Query Match 1.7%; Score 7; DB 1; Length 281;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 PGAGSGA 275  
Db 46 PGAGSGA 52  
|||||

RESULT 39  
US-09-252-991A-26761  
; Sequence 26761, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26761  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26761

Query Match 1.7%; Score 7; DB 4; Length 281;

Best Local Similarity 100.0%; Pred. No. 1.8e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;

QY 102 PVAPAAS 108  
Db 61 PVAPAAS 67

## RESULT 40

US-08-397-633A-76  
; Sequence 76, Application US/08397633A  
; Patent No. 5773577

## ; GENERAL INFORMATION:

; APPLICANT: Cappello, Joseph  
; TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE  
; TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING  
; NUMBER OF SEQUENCES: 105  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBRACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/397,633A  
; FILING DATE:

## ; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Bertram I  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 76:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 287 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-397-633A-76

Query Match 1.7%; Score 7; DB 1; Length 287;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 PGAGSGA 275  
Db 49 PGAGSGA 55

## RESULT 41

US-09-252-991A-22293  
; Sequence 22293, Application US/09252991A  
; Patent No. 6551795

## ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22293  
; LENGTH: 296  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22293

Query Match 1.7%; Score 7; DB 4; Length 296;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 PATLAQP 99  
Db 265 PATLAQP 271

## RESULT 42

US-09-495-406-17  
; Sequence 17, Application US/09495406  
; Patent No. 6503744

## ; GENERAL INFORMATION:

; APPLICANT: Gilbert, Michel  
; APPLICANT: Wakarchuk, Warren W.  
; APPLICANT: National Research Council of Canada  
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics  
; FILE REFERENCE: 019633-000110US  
; CURRENT APPLICATION NUMBER: US/09/495,406  
; CURRENT FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: US 60/118,213  
; PRIOR FILING DATE: 1999-02-01  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 303  
; TYPE: PRT  
; ORGANISM: Campylobacter jejuni  
US-09-495-406-17

Query Match 1.7%; Score 7; DB 4; Length 303;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 ILDMGGG 173  
Db 106 ILDMGGG 112

## RESULT 43

US-09-134-001C-3568  
; Sequence 3568, Application US/09134001C  
; Patent No. 6380370

## ; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3568  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3568

Query Match 1.7%; Score 7; DB 4; Length 341;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;

```
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 KLVNKTLL 8
    |||||
Db 211 KLVNKTLL 217

RESULT 44
US-09-252-991A-24458
; Sequence 24458, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24458
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24458

Query Match 1.7%; Score 7; DB 4; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 344 LAAAVPQ 350
    |||||
Db 228 LAAAVPQ 234

RESULT 45
US-09-107-532A-6666
; Sequence 6666, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS PRECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
```

```
TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6666:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...360
; SEQUENCE DESCRIPTION: SEQ ID NO: 6666:
US-09-107-532A-6666

Query Match 1.7%; Score 7; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 134 PAATVAA 140
    |||||
Db 309 PAATVAA 315

RESULT 46
US-09-252-991A-26295
; Sequence 26295, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26295
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (244)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-26295

Query Match 1.7%; Score 7; DB 4; Length 364;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 104 APAASVA 110
    |||||
Db 291 APAASVA 297

RESULT 47
US-09-252-991A-23727
; Sequence 23727, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
```

; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23727  
; LENGTH: 382  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23727

Query Match 1.7%; Score 7; DB 4; Length 382;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 AQPVARA 214  
Db 109 AQPVARA 115  
|||||

RESULT 48  
US-09-377-557-18  
; Sequence 18, Application US/09377557  
; Patent No. 6297055  
; GENERAL INFORMATION:  
; APPLICANT: Falco, S. Carl  
; APPLICANT: Farnodu, Layo O.  
; APPLICANT: Orozco, Emil M. Jr.  
; TITLE OF INVENTION: Amino Acid Decarboxylases  
; FILE REFERENCE: BB-1237  
; CURRENT APPLICATION NUMBER: US/09/377,557  
; CURRENT FILING DATE: 1999-08-19  
; EARLIER APPLICATION NUMBER: 60/099,493  
; EARLIER FILING DATE: September 8, 1998  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 18  
; LENGTH: 391  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-09-377-557-18

Query Match 1.7%; Score 7; DB 3; Length 391;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 QAAPAAAT 137  
Db 25 QAAPAAAT 31  
|||||

RESULT 49  
US-09-252-991A-21075  
; Sequence 21075, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21075  
; LENGTH: 406  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21075

Query Match 1.7%; Score 7; DB 4; Length 406;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 138 VAATDDA 144  
Db 185 VAATDDA 191  
|||||

RESULT 50  
US-09-252-991A-27669  
; Sequence 27669, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27669  
; LENGTH: 415  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27669

Query Match 1.7%; Score 7; DB 4; Length 415;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 PGAGSGA 275  
Db 120 PGAGSGA 126  
|||||

RESULT 51  
US-09-252-991A-22253  
; Sequence 22253, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22253  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22253

Query Match 1.7%; Score 7; DB 4; Length 420;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 GOLAAAV 348  
Db 213 GOLAAAV 219  
|||||

RESULT 52  
US-09-252-991A-23618  
; Sequence 23618, Application US/09252991A  
; Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIORITY FILING DATE: 1999-02-18  
PRIORITY APPLICATION NUMBER: US 60/074,788  
PRIORITY FILING DATE: 1998-02-18  
PRIORITY APPLICATION NUMBER: US 60/094,190  
PRIORITY FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 23618  
LENGTH: 423  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23618

Query Match 1.7%; Score 7; DB 4; Length 423;

Best Local Similarity 100.0%; Pred. No. 2.6e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;

Qy 343 QLAAPV 349  
Db 212 QLAAPV 218

RESULT 53  
US-08-476-509B-2  
Sequence 2, Application US/08476509B  
Patent No. 6034212  
GENERAL INFORMATION:  
APPLICANT: SUDOL, MARIUS  
APPLICANT: PEER, BORK  
APPLICANT: HENRY, CHEN  
TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A  
TITLE OF INVENTION: SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE  
TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTICS  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,509B  
FILING DATE: 01-DEC-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-101 CIP  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 448 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-476-509B-2

Query Match 1.7%; Score 7; DB 3; Length 448;

Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 PGAGSGA 275  
Db 28 PGAGSGA 34

RESULT 54  
US-08-948-997-6  
Sequence 6, Application US/08948997  
Patent No. 6008020  
GENERAL INFORMATION:  
APPLICANT: HASTINGS, GREGG  
APPLICANT: COLEMAN, TIM  
TITLE OF INVENTION: BRAIN-ASSOCIATED INHIBITOR OF  
TITLE OF INVENTION: TISSUE-TYPE PLASMINOGEN ACTIVATOR  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
STREET: 9410 KEY WEST AVENUE  
CITY: ROCKVILLE  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/948,997  
FILING DATE: Oct-10-97  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: A. ANDERS BROOKES  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF336  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 465 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-948-997-6

Query Match 1.7%; Score 7; DB 3; Length 465;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 PGAGSGA 275  
Db 4 PGAGSGA 10

RESULT 55  
US-09-348-817A-6  
Sequence 6, Application US/09348817A  
Patent No. 6191260  
GENERAL INFORMATION:  
APPLICANT: Hastings et al.  
TITLE OF INVENTION: Brain-Associated Inhibitor of Tissue-Type Plasminogen  
TITLE OF INVENTION: Activator  
FILE REFERENCE: PF336D1  
CURRENT APPLICATION NUMBER: US/09/348,817A  
CURRENT FILING DATE: 1999-07-08  
PRIORITY APPLICATION NUMBER: 08/948,997  
PRIORITY FILING DATE: 1997-10-10

;; PRIOR APPLICATION NUMBER: 60/028,117  
;; PRIOR FILING DATE: 1996-10-11  
;; NUMBER OF SEQ ID NOS: 17  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 6  
;; LENGTH: 465  
;; TYPE: PRT  
;; ORGANISM: Mus musculus  
US-09-348-817A-6

Query Match 1.7%; Score 7; DB 3; Length 465;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 PGAGSGA 275  
DB 4 PGAGSGA 10

RESULT 56  
US-09-722-292-6  
; Sequence 6, Application US/09722292  
; Patent No. 6541452  
; GENERAL INFORMATION:  
; APPLICANT: Hastings et al.  
; TITLE OF INVENTION: Brain-Associated Inhibitor of Tissue-Type Plasminogen  
; TITLE OF INVENTION: Activator  
; FILE REFERENCE: PF336D1  
; CURRENT APPLICATION NUMBER: US/09/722,292  
; CURRENT FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 09/348,817  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: 60/028,117  
; PRIOR FILING DATE: 1996-10-11  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-722-292-6

Query Match 1.7%; Score 7; DB 4; Length 465;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 PGAGSGA 275  
DB 4 PGAGSGA 10

RESULT 57  
US-09-252-991A-23144  
; Sequence 23144, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23144  
; LENGTH: 479  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23144

Query Match 1.7%; Score 7; DB 4; Length 479;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 QPAAAPVA 104  
DB 302 QPAAAPVA 308

RESULT 58  
US-08-348-518C-2  
; Sequence 2, Application US/08348518C  
; Patent No. 6022740  
; GENERAL INFORMATION:  
; APPLICANT: SUDOL, MARIUS  
; APPLICANT: PEER, BORK  
; APPLICANT: HENRY, CHEN  
; TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A  
; TITLE OF INVENTION: SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE  
; TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/348,518C  
; FILING DATE: 01-DEC-1994  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-101  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 486 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-348-518C-2

Query Match 1.7%; Score 7; DB 3; Length 486;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 PGAGSGA 275  
DB 28 PGAGSGA 34

RESULT 59  
US-09-252-991A-27501  
; Sequence 27501, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27501  
; LENGTH: 495  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27501

Query Match 1.7%; Score 7; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 AKASSG 83  
Db 458 AKASSG 464

## RESULT 60

US-09-252-991A-16737  
; Sequence 16737, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 16737  
; LENGTH: 519  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-16737

Query Match 1.7%; Score 7; DB 4; Length 519;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 213 RAPAAQG 219  
Db 2 RAPAAQG 8

## RESULT 61

US-09-252-991A-22537  
; Sequence 22537, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22537  
; LENGTH: 534  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22537

Query Match 1.7%; Score 7; DB 4; Length 534;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 AAPVAPA 106  
Db 267 AAPVAPA 273

## RESULT 62

US-09-252-991A-26327  
; Sequence 26327, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26327  
; LENGTH: 549  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26327

Query Match 1.7%; Score 7; DB 4; Length 549;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 AASNLVF 156  
Db 40 AASNLVF 46

## RESULT 63

US-09-252-991A-23080  
; Sequence 23080, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23080  
; LENGTH: 561  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23080

Query Match 1.7%; Score 7; DB 4; Length 561;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 PVQPSA 242  
Db 39 PVQPSA 45

## RESULT 64



US-09-205-258-573  
; Sequence 573, Application US/09205258  
; Patent No. 6525174  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 207 Human Secreted Proteins  
; FILE REFERENCE: PZ007P1  
; CURRENT APPLICATION NUMBER: US/09/205,258  
; CURRENT FILING DATE: 1998-12-04  
; EARLIER APPLICATION NUMBER: PCT/US98/11422  
; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,885  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,375  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,881  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,880  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,896  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,020  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,876  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,895  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,884  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,894  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,971  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,964  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,882  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,899  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,893  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,900  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,901  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,892  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,915  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,019  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,970  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,972  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,916  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,373  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,875  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,374  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,917  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,949  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,974  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,883  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,897  
; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,898  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,962  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,963  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,877  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,878  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/070,923  
; EARLIER FILING DATE: 1997-12-18  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/094,657  
; EARLIER FILING DATE: 1998-07-30  
; NUMBER OF SEQ ID NOS: 1227  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 573  
; LENGTH: 567  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURES:  
; NAME/KEY: SITE  
; LOCATION: (409)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-205-258-573

Query Match 1.7%; Score 7; DB 4; Length 567;  
Best Local Similarity 100.0%; Pred.No. 3.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 GASTATT 89  
|||  
DB 79 GASTATT 85

RESULT 65  
US-09-199-290-7  
; Sequence 7, Application US/09199290  
; Patent No. 6235084  
; GENERAL INFORMATION:  
; APPLICANT: Nielsen, Bjarne R.  
; APPLICANT: Nielsen, Ruby  
; APPLICANT: Lehmebeck, Jan  
; TITLE OF INVENTION: Thermostable Glucoamylase  
; FILE REFERENCE: 5279.200-US  
; CURRENT APPLICATION NUMBER: US/09/199,290  
; CURRENT FILING DATE: 1998-11-24  
; EARLIER APPLICATION NUMBER: 1557/97  
; EARLIER FILING DATE: 1997-12-30  
; EARLIER APPLICATION NUMBER: 0925/98  
; EARLIER FILING DATE: 1998-07-10  
; EARLIER APPLICATION NUMBER: 60/070,746  
; EARLIER FILING DATE: 1998-01-08  
; EARLIER APPLICATION NUMBER: 60/094,344  
; EARLIER FILING DATE: 1998-07-28  
; EARLIER APPLICATION NUMBER: 08/979,673  
; EARLIER FILING DATE: 1997-11-26  
; EARLIER APPLICATION NUMBER: 09/107,657  
; EARLIER FILING DATE: 1998-06-30  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 591  
; TYPE: PRT  
; ORGANISM: Talaromyces emersonii  
US-09-199-290-7

Query Match 1.7%; Score 7; DB 3; Length 591;  
Best Local Similarity 100.0%; Pred.No. 3.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 160 LEQTIQ 166  
|||||||  
Db 76 LEQTIQ 82

RESULT 66  
US-09-252-991A-23472  
; Sequence 23472, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23472  
; LENGTH: 617  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23472

Query Match 1.7%; Score 7; DB 4; Length 617;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 275 ALDALRQ 281  
|||||||  
Db 341 ALDALRQ 347

RESULT 67  
US-09-199-290-34  
; Sequence 34, Application US/09199290  
; Patent No. 6255084  
; GENERAL INFORMATION:  
; APPLICANT: Nielsen, Bjarne R.  
; APPLICANT: Nielsen, Ruby  
; APPLICANT: Lehmebeck, Jan  
; TITLE OF INVENTION: Thermostable Glucoamylase  
; FILE REFERENCE: 5279.200-US  
; CURRENT APPLICATION NUMBER: US/09/199,290  
; CURRENT FILING DATE: 1998-11-24  
; EARLIER APPLICATION NUMBER: 1557/97  
; EARLIER FILING DATE: 1997-12-30  
; EARLIER APPLICATION NUMBER: 0925/98  
; EARLIER FILING DATE: 1998-07-10  
; EARLIER APPLICATION NUMBER: 60/070,746  
; EARLIER FILING DATE: 1998-01-08  
; EARLIER APPLICATION NUMBER: 60/094,344  
; EARLIER FILING DATE: 1998-07-28  
; EARLIER APPLICATION NUMBER: 08/979,673  
; EARLIER FILING DATE: 1997-11-26  
; EARLIER APPLICATION NUMBER: 09/107,657  
; EARLIER FILING DATE: 1998-06-30  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 34  
; LENGTH: 618  
; TYPE: PRT  
; ORGANISM: Talaromyces emersonii  
US-09-199-290-34

Query Match 1.7%; Score 7; DB 3; Length 618;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 160 LEQTIQ 166

Db 103 LEQTIQ 109  
|||||||

RESULT 68  
US-09-252-991A-21585  
; Sequence 21585, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21585  
; LENGTH: 619  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21585

Query Match 1.7%; Score 7; DB 4; Length 619;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 138 VAATDDA 144  
|||||||  
Db 589 VAATDDA 595

RESULT 69  
US-09-252-991A-28001  
; Sequence 28001, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28001  
; LENGTH: 657  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28001

Query Match 1.7%; Score 7; DB 4; Length 657;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 97 AQPAPV 103  
|||||||  
Db 538 AQPAPV 544

RESULT 70  
US-08-795-868-14  
; Sequence 14, Application US/08795868  
; Patent No. 5846773  
; GENERAL INFORMATION:  
; APPLICANT: Lee, Mu-En  
; APPLICANT: Hsieh, Chung-Ming

;/ TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC  
;/ TITLE OF INVENTION: AND STRIATED-SPECIFIC MUSCLE CELL ISOFORMS AND USES THEREOF  
;/ NUMBER OF SEQUENCES: 19  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: Fish & Richardson, P.C.  
;/ STREET: 225 Franklin Street  
;/ CITY: Boston  
;/ STATE: MA  
;/ COUNTRY: US  
;/ ZIP: 02110-2804  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Diskette  
;/ COMPUTER: IBM Compatible  
;/ OPERATING SYSTEM: DOS  
;/ SOFTWARE: FastSeq for Windows Version 2.0  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/795,868  
;/ FILING DATE: 06-FEB-1997  
;/ CLASSIFICATION: 424  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: 08/494,577  
;/ FILING DATE: 22-JUN-1995  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Fraser, Janis K.  
;/ REGISTRATION NUMBER: 34,819  
;/ REFERENCE/DOCKET NUMBER: 05433/032001  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: 617-542-5070  
;/ TELEFAX: 617-542-8906  
;/ INFORMATION FOR SEQ ID NO: 14:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 661 amino acids  
;/ TYPE: amino acid  
;/ STRANDEDNESS: not relevant  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: protein  
;/ FRAGMENT TYPE: Internal  
;/ US-08-795-868-14

Query Match 1.7%; Score 7; DB 2; Length 661;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 346 AAVPQTL 352  
Db 350 AAVPQTL 356

RESULT 71  
US-09-303-069-14  
;/ Sequence 14, Application US/09303069A  
;/ Patent No. 6350592  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Lee, Mu-En  
;/ APPLICANT: Hsieh, Chung-Ming  
;/ TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC  
;/ TITLE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF  
;/ FILE REFERENCE: 05433/039001  
;/ CURRENT APPLICATION NUMBER: US/09/303,069A  
;/ CURRENT FILING DATE: 1999-04-30  
;/ EARLIER APPLICATION NUMBER: US 09/134,250  
;/ EARLIER FILING DATE: 1998-08-14  
;/ NUMBER OF SEQ ID NOS: 24  
;/ SOFTWARE: FastSeq for Windows Version 3.0  
;/ SEQ ID NO 14  
;/ LENGTH: 661  
;/ TYPE: PRT  
;/ ORGANISM: Homo sapiens  
;/ US-09-303-069-14

Query Match 1.7%; Score 7; DB 4; Length 661;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 346 AAVPQTL 352  
Db 350 AAVPQTL 356

RESULT 72  
US-09-134-250-14  
;/ Sequence 14, Application US/09134250B  
;/ Patent No. 6399753  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Lee, Mu-En  
;/ APPLICANT: Hsieh, Chung-Ming  
;/ TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC  
;/ TITLE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF  
;/ FILE REFERENCE: 05433/038001  
;/ CURRENT APPLICATION NUMBER: US/09/134,250B  
;/ CURRENT FILING DATE: 1998-08-14  
;/ EARLIER APPLICATION NUMBER: US 08/795,868  
;/ EARLIER FILING DATE: 1997-02-06  
;/ EARLIER APPLICATION NUMBER: US 08/494,577  
;/ EARLIER FILING DATE: 1995-06-22  
;/ NUMBER OF SEQ ID NOS: 20  
;/ SOFTWARE: FastSeq for Windows Version 3.0  
;/ SEQ ID NO 14  
;/ LENGTH: 661  
;/ TYPE: PRT  
;/ ORGANISM: Homo sapiens  
;/ US-09-134-250-14

Query Match 1.7%; Score 7; DB 4; Length 661;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 346 AAVPQTL 352  
Db 350 AAVPQTL 356

RESULT 73  
US-09-252-991A-20752  
;/ Sequence 20752, Application US/09252991A  
;/ Patent No. 6551795  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Marc J. Rubenfield et al.  
;/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
;/ FILE REFERENCE: 107196.136  
;/ CURRENT APPLICATION NUMBER: US/09/252,991A  
;/ CURRENT FILING DATE: 1999-02-18  
;/ PRIOR APPLICATION NUMBER: US 60/074,788  
;/ PRIOR FILING DATE: 1998-02-18  
;/ PRIOR APPLICATION NUMBER: US 60/094,190  
;/ PRIOR FILING DATE: 1998-07-27  
;/ NUMBER OF SEQ ID NOS: 33142  
;/ SEQ ID NO 20752  
;/ LENGTH: 749  
;/ TYPE: PRT  
;/ ORGANISM: Pseudomonas aeruginosa  
;/ US-09-252-991A-20752

Query Match 1.7%; Score 7; DB 4; Length 749;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 275 ALDALRQ 281  
Db 446 ALDALRQ 452

RESULT 74  
US-09-252-991A-23774  
;/ Sequence 23774, Application US/09252991A

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23774
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23774

Query Match      1.7%; Score 7; DB 4; Length 798;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      104 APAASVA 110
      |||||
Db      243 APAASVA 249

RESULT 75
US-09-252-991A-25050
; Sequence 25050, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25050
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25050

Query Match      1.7%; Score 7; DB 4; Length 802;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      212 ARAPAAG 218
      |||||
Db      525 ARAPAAG 531

Search completed: December 17, 2003, 06:36:28
Job time : 17.1462 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:27:31 ; Search time 16.6624 Seconds  
(without alignments)  
2123.954 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 368

Sequence: 1 MKLTVKTLKTHFEIRVQPN.....CORNEELAANYLLEHAGEED 368

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : PIR\_76.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	5.7	367	2	F96827	protein F20B17.8 [
2	15	4.1	113	2	H86296	P309.1 protein - A
3	15	4.1	279	2	T14337	RAD23 protein, iso
4	15	4.1	382	2	T14336	RAD23 protein, iso
5	14	3.8	246	2	G86296	T24D18.27 protein
6	13	3.5	392	2	T04150	RAD23 protein homo
7	9	2.4	225	2	H75381	DNA-binding respon
8	9	2.4	368	2	T40115	uv excision repair
9	9	2.4	753	2	A96747	probable RNA-bindi
10	8	2.2	111	2	T14306	glycine-rich prote
11	8	2.2	173	2	JQ1064	glycine-rich prote
12	8	2.2	180	2	T49530	related to glycine
13	8	2.2	210	2	JQ1060	glycine-rich prote
14	8	2.2	220	2	T14441	glycine-rich prote
15	8	2.2	276	2	B25345	tropomycin, cardia
16	8	2.2	338	1	KNMU	glycine-rich cell
17	8	2.2	384	1	A26099	glycine-rich cell
18	8	2.2	401	2	C96614	hypothetical prote
19	8	2.2	465	1	S01820	glycine-rich cell
20	8	2.2	548	2	T28910	hypothetical prote
21	8	2.2	631	2	G64874	probable membrane
22	8	2.2	712	2	A45638	immunodominant mic
23	8	2.2	809	2	S55344	outer envelope mem
24	8	2.2	823	2	T02812	probable membrane
25	8	2.2	1010	2	I40329	brkA proteoin - Bor
26	8	2.2	1454	2	T13709	diacylglycerol kin
27	8	2.2	1718	2	T14603	hypothetical prote
28	8	2.2	2425	2	D69426	surface layer prot
29	8	2.2	3016	2	S77300	hypothetical prote

30	7	1.9	42	2	A71429	hypothetical prote
31	7	1.9	50	2	B49410	t-complex polypept
32	7	1.9	76	2	A39034	neurogranin - bovi
33	7	1.9	78	2	A57288	neurogranin - rat
34	7	1.9	78	2	I47043	neurogranin RC3 [i
35	7	1.9	94	2	S49470	hypothetical prote
36	7	1.9	96	2	F97355	uncharacterized co
37	7	1.9	102	2	S69752	hypothetical prote
38	7	1.9	102	2	T45500	hypothetical prote
39	7	1.9	108	2	G86252	hypothetical prote
40	7	1.9	112	2	C55993	hypothetical prote
41	7	1.9	119	2	S58445	lipoprotein D - Sa
42	7	1.9	119	2	D83723	hypothetical prote
43	7	1.9	121	2	A24272	lg heavy chain pre
44	7	1.9	172	2	T36107	probable serine/ar
45	7	1.9	180	2	C82243	conserved hypotet
46	7	1.9	182	2	I59203	gene HOX2.8 protei
47	7	1.9	195	2	AF0830	probable membrane
48	7	1.9	201	2	T00799	hypothetical prote
49	7	1.9	201	2	G01204	twist protein homo
50	7	1.9	206	2	I53066	gene M-twist prote
51	7	1.9	207	2	T07854	germin-like protei
52	7	1.9	209	2	H69037	conserved hypotet
53	7	1.9	215	2	C87542	hypothetical prote
54	7	1.9	216	2	T48482	hypothetical prote
55	7	1.9	218	2	AF3273	ribosomal protein
56	7	1.9	224	2	A53143	testis-determining
57	7	1.9	228	2	A98166	hypothetical prote
58	7	1.9	228	2	AF3121	hypothetical prote
59	7	1.9	231	2	T07358	ribosomal protein
60	7	1.9	234	2	G95989	hypothetical glyci
61	7	1.9	239	2	I46082	CD8 alpha-chain -
62	7	1.9	247	2	T46968	flavoprotein [impo
63	7	1.9	248	2	T33230	hypothetical prote
64	7	1.9	249	2	B33144	homeotic protein U
65	7	1.9	255	2	B84777	hypothetical prote
66	7	1.9	266	1	CRBL	calpain [SC 3.4.22
67	7	1.9	268	2	S09860	hypothetical prote
68	7	1.9	272	1	G83023	3',5'-cyclic-nucle
69	7	1.9	274	2	A83583	probable biotin sy
70	7	1.9	276	2	AH2885	hypothetical prote
71	7	1.9	276	2	E97651	dipeptide transpor
72	7	1.9	284	2	S74256	homeotic protein s
73	7	1.9	290	2	T23416	hypothetical prote
74	7	1.9	291	1	S31415	glycine-rich prote
75	7	1.9	309	2	T00503	probable MYB fami
76	7	1.9	320	2	F82763	D-alanine-D-alanin
77	7	1.9	323	2	S16318	homeotic protein H
78	7	1.9	325	2	A75415	homoserine dehydro
79	7	1.9	327	2	B75341	probable oxidoredu
80	7	1.9	329	2	A37864	SCL protein homo
81	7	1.9	330	2	S74255	homeotic protein s
82	7	1.9	333	2	AF0407	lipoprotein [impor
83	7	1.9	339	2	F97190	phenylalanyl-tRNA
84	7	1.9	341	2	AG1210	glucitol dehydroge
85	7	1.9	342	2	S14432	heterogeneous ribo
86	7	1.9	344	2	T33057	hypothetical prote
87	7	1.9	344	2	T48827	hypothetical prote
88	7	1.9	346	2	B98222	probable transcrip
89	7	1.9	347	2	AF3064	transcription regu
90	7	1.9	348	2	AH0090	probable regulator
91	7	1.9	362	2	T29278	hypothetical prote
92	7	1.9	366	2	T44827	probable outer mem
93	7	1.9	366	2	T27309	hypothetical prote
94	7	1.9	367	2	JC6087	helix-loop-helix t
95	7	1.9	373	2	AH0855	lipoprotein NlpD p
96	7	1.9	374	2	T33173	hypothetical prote
97	7	1.9	375	2	E96567	hypothetical prote
98	7	1.9	376	2	AD1621	heat shock protein
99	7	1.9	377	2	T43739	heat shock protein
100	7	1.9	377	2	AH1258	heat shock protein

## ALIGNMENTS

## RESULT 1

F96827  
 Protein F20B17.8 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 17-May-2002  
 C/Accession: F96827  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: F96827  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-367 <STO>  
 A:Cross-references: GB:AE005173; NID:g7715605; PIDN:AAF68123.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F20B17.8  
 A:Map position: 1  
 C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 5.7%; Score 21; DB 2; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-13;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 VIEAFLACDRNEELAAAYLLE 362  
 |||||  
 Db 340 VIEAFLACDRNEELAAAYLLE 360

## RESULT 2

H86296  
 F309.1 protein - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001  
 C/Accession: H86296  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: H86296  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-113 <STO>  
 A:Cross-references: GB:AE005173; NID:g4966345; PIDN:AAD34676.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match 4.1%; Score 15; DB 2; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 CDRNEELAAAYLLEH 363  
 |||||  
 Db 93 CDRNEELAAAYLLEH 107

## RESULT 3

T14337  
 RAD23 protein, isoform II - carrot  
 C:Species: Daucus carota (carrot)  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
 C/Accession: T14337  
 R:Sturm, A.; Leinhardt, S.  
 Plant J. 13, 815-821, 1998  
 A>Title: Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in yeast.  
 A:Reference number: Z17989; MUID:98345997; PMID:9681019  
 A:Accession: T14337  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-379 <STU>  
 A:Cross-references: EMBL:Y12014; NID:g1914684; PIDN:CAA72742.1; PID:g1914685  
 A:Experimental source: subspecies Queen Anne's lace, isolate W001C  
 C:Genetics:  
 A:Gene: RAD23-2  
 C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 4.1%; Score 15; DB 2; Length 379;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 VIEAFLACDRNEELA 356  
 |||||  
 Db 352 VIEAFLACDRNEELA 366

## RESULT 4

T14336  
 RAD23 protein, isoform I - carrot  
 C:Species: Daucus carota (carrot)  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
 C/Accession: T14336  
 R:Sturm, A.; Leinhardt, S.  
 Plant J. 13, 815-821, 1998  
 A>Title: Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in yeast.  
 A:Reference number: Z17989; MUID:98345997; PMID:9681019  
 A:Accession: T14336  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-382 <STU>  
 A:Cross-references: EMBL:Y12013; NID:g1914682; PIDN:CAA72741.1; PID:g1914683  
 A:Experimental source: subspecies Queen Anne's lace, isolate W001C  
 C:Genetics:  
 A:Gene: RAD23-1  
 C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 4.1%; Score 15; DB 2; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 7.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 NNPERAVELYSGIP 189  
 |||||  
 Db 187 NNPERAVELYSGIP 201

## RESULT 5

G86296  
 T24D18.27 protein - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001  
 C/Accession: G86296  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.; ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: G86296  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-246 <STO>  
 A:Cross-references: GB:AE005172; NID:G6587822; PIDN:AAF18513.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match 3.8%; Score 14; DB 2; Length 246;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-06; Mismatches 0; Indels 0; Gaps 0;  
 Matches 14; Conservative 0;

QY 167 RALRAAYNNPERAV 180  
 |||||  
 DB 155 RALRAAYNNPERAV 168

RESULT 6  
 T04150  
 RAD23 protein homolog - rice  
 C:Species: Oryza sativa (rice)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
 C:Accession: T04150  
 R:Schultz, T.F.; Quatrano, R.S.  
 Plant Mol. Biol. 34, 557-562, 1997  
 A>Title: Characterization and expression of a rice RAD23 gene.  
 A:Reference number: Z08695; MUID:97369378; PMID:9225866  
 A:Accession: T04150  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-392 <SCH>  
 A:Cross-references: EMBL:U63530; NID:g1488296; PIDN:AAB65841.1; PID:g1488297  
 A:Experimental source: cv. Nipponbare  
 C:Genetics:  
 A:Gene: RAD23  
 C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 3.5%; Score 13; DB 2; Length 392;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-05; Mismatches 0; Indels 0; Gaps 0;  
 Matches 13; Conservative 0;

QY 174 NNPERAVEYLYSG 186  
 |||||  
 DB 187 NNPERAVEYLYSG 199

RESULT 7  
 H75381  
 DNA-binding response regulator - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 12-May-2003  
 C:Accession: H75381  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: H75381  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-225 <WHI>  
 A:Cross-references: GB:AE001999; GB:AE000513; NID:g6459316; PIDN:AAF11120.1; PID:g645932  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR1558  
 A:Map position: 1  
 C:Superfamily: response regulator with an HTH DNA-binding domain, NarL type; response re

Query Match 2.4%; Score 9; DB 2; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 0.59;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 57 STLEENKVN 65  
 |||||  
 DB 95 STLEENKVN 103

RESULT 8  
 T40115  
 uv excision repair protein rad23 homolog - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 02-Sep-2000  
 C:Accession: T40115; T51298  
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Taylor, K.; Harris, D.  
 submitted to the EMBL Data Library, September 1998  
 A:Reference number: Z21906  
 A:Accession: T40115  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-368 <WOO>  
 A:Cross-references: EMBL:AL031788; PIDN:CAA21170.1; GSPDB:GN00067; SPDB:SPBC2D10.12  
 A:Experimental source: strain 972h-; cosmid c2D10  
 R:Zhao, Y.; Elder, R.T.  
 submitted to the EMBL Data Library, July 1999  
 A:Description: A fission yeast orthologue (rhp23) of the human nucleotide excision rep  
 A:Reference number: Z25362  
 A:Accession: T51298  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-368 <ZHA>  
 A:Cross-references: EMBL:AF174293; PIDN:AAD51975.1  
 A:Experimental source: strain SP223  
 C:Genetics:  
 A:Gene: rhp23; SPDB:SPBC2D10.12  
 A:Map position: 2  
 A:Introns: 23/3; 48/1; 328/3; 351/3  
 C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 2.4%; Score 9; DB 2; Length 368;  
 Best Local Similarity 100.0%; Pred. No. 0.93; Mismatches 0; Indels 0; Gaps 0;  
 Matches 9; Conservative 0;

QY 352 NEELAANYL 360  
 |||||  
 DB 349 NEELAANYL 357

RESULT 9  
 A96747  
 Probable RNA-binding domain T10D10.21 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: A96747  
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alons  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 815-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: A96747  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-753 <STO>  
 A:Cross-references: GB:AE005173; NID:g6730769; PIDN:AAF27158.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: T10D10.21  
 A:Map position: 1

Query Match 2.4%; Score 9; DB 2; Length 753;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 340 ARVIEAFLA 348  
|||||  
Db 542 ARVIEAFLA 550

RESULT 10  
T14306  
glycine-rich protein - carrot (fragment)  
C:Species: Daucus carota (carrot)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T14306  
R:Lin, X.; Hwang, G.J.; Zimmerman, J.L.  
submitted to the EMBL Data Library, January 1996  
A:Description: Isolation and characterization of a diverse set of genes from carrot som

A:Reference number: Z17968  
A:Accession: T14306  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-111 <LIN>  
A:Cross-references: EMBL:U47097; NID:g1276970; PID:g1276971  
A:Experimental source: strain Danver Half-long

Query Match 2.2%; Score 8; DB 2; Length 111;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||  
Db 18 AGGAGGG 25

RESULT 11  
JQ1064  
glycine-rich protein 5 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 20-Aug-1999  
C:Accession: JQ1064  
R:de Oliveira, D.E.; Seurinck, J.; Inze, D.; Van Montagu, M.; Botterman, J.  
Plant Cell 2, 427-436, 1990  
A:Title: Differential expression of five Arabidopsis genes encoding glycine-rich protein  
A:Reference number: JQ1060; MUID:93044485; PMID:2152168  
A:Molecule type: mRNA  
A:Residues: 1-173 <DEO>  
A:Cross-references: GB:S47414; NID:g259450; PIDN:AAB24077.1; PID:g259451  
A:Experimental source: strain C24  
C:Superfamily: glycine-rich cell wall structural protein 1

Query Match 2.2%; Score 8; DB 2; Length 173;  
Best Local Similarity 100.0%; Pred. No. 4.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||  
Db 84 AGGAGGG 91

RESULT 12  
T49530  
related to glycine-rich cell wall structural protein [imported] - Neurospora crassa  
N:Alternate names: protein B21J21.90  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49530  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49530  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-180 <SCH>  
A:Cross-references: EMBL:AL355929; GSPDB:GN00116; NCSP:B21J21.90  
A:Experimental source: BAC clone B21J21; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B21J21.90  
A:Map position: 6

Query Match 2.2%; Score 8; DB 2; Length 180;  
Best Local Similarity 100.0%; Pred. No. 5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||  
Db 60 AGGAGGG 67

RESULT 13  
JQ1060

glycine-rich protein 1 - Arabidopsis thaliana (fragment)  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 20-Aug-1999  
C:Accession: JQ1060  
R:de Oliveira, D.E.; Seurinck, J.; Inze, D.; Van Montagu, M.; Botterman, J.  
Plant Cell 2, 427-436, 1990  
A:Title: Differential expression of five Arabidopsis genes encoding glycine-rich prote  
A:Reference number: JQ1060; MUID:93044485; PMID:2152168  
A:Molecule type: mRNA  
A:Residues: 1-210 <DEO>  
A:Cross-references: GB:S47405; NID:g259442; PIDN:AAB24073.1; PID:g259443  
A:Experimental source: strain C24  
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 2.2%; Score 8; DB 2; Length 210;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||  
Db 148 AGGAGGG 155

RESULT 14  
T14441

glycine-rich protein - wild cabbage (fragment)  
C:Species: Brassica oleracea (wild cabbage)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Jun-2000  
C:Accession: T14441  
R:Booker, J.P.; Crov, R.D.  
submitted to the EMBL Data Library, July 1996  
A:Reference number: Z18090  
A:Accession: T14441  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-220 <BOO>  
A:Cross-references: EMBL:Z74892  
A:Experimental source: stigma  
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 2.2%; Score 8; DB 2; Length 220;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||  
Db 16 AGGAGGG 23

RESULT 15  
B25345

C:Species: Oryctolagus cuniculus (domestic rabbit)  
troponin T, cardiac muscle, minor isoform - rabbit



C;Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 23-Feb-1997

C;Accession: B25345

R;Pearlatone, J.R.; Carpenter, M.R.; Smillie, L.B.

J. Biol. Chem. 261, 16795-16810, 1986

A;Title: Amino acid sequence of rabbit cardiac troponin T.

A;Reference number: A92569; MUID:87057385; PMID:3782144

A;Accession: B25345

A;Molecule type: protein

A;Residues: 1-276 <PEA>

C;Superfamily: troponin T

C;Keywords: alternative splicing; cardiac muscle; differentiation; heart; muscle; phosph

Query Match 2.2%; Score 8; DB 2; Length 276;

Best Local Similarity 100.0%; Pred. No. 7.5;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241

Db 21 AGGGAGGG 28

RESULT 16

KNMU

glycine-rich cell wall protein precursor - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Jul-1999

C;Accession: S17732

R;Quigley, F.; Villiot, M.L.; Mache, R.

Plant Mol. Biol. 17, 949-952, 1991

A;Title: Nucleotide sequence and expression of a novel glycine-rich protein gene from A

A;Reference number: S17732; MUID:92003708; PMID:1912511

A;Accession: S17732

A;Molecule type: DNA

A;Residues: 1-338 <QUI>

A;Cross-references: EMBL:X58338; NID:gl6292; PIDN:CAA41249.1; PID:gl6293

C;Genetics:

A;Introns: 203/1

C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

C;Keywords: cell wall; structural protein

F;1-20/Domain: signal sequence #status predicted <SIG>

F;21-338/Product: glycine-rich cell wall protein #status predicted <MAT>

Query Match 2.2%; Score 8; DB 1; Length 338;

Best Local Similarity 100.0%; Pred. No. 9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241

Db 52 AGGGAGGG 59

RESULT 17

A26099

glycine-rich cell wall structural protein - garden petunia

C;Species: Petunia x hybrida (garden petunia)

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C;Accession: A26099

R;Condit, C.M.; Meagher, R.B.

Nature 323, 178-181, 1986

A;Title: A gene encoding a novel glycine-rich structural protein of petunia.

A;Reference number: A26099

A;Accession: A26099

A;Molecule type: mRNA

A;Residues: 1-384 <CON>

C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 2.2%; Score 8; DB 1; Length 384;

Best Local Similarity 100.0%; Pred. No. 10;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241

Db 75 AGGGAGGG 82

RESULT 20

T28910

hypothetical protein T26C11.4 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000

RESULT 18

C96614

hypothetical protein T18T24.2 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: C96614

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alons

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K

ansen, N.P.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.B.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: C96614

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-401 <STO>

A;Cross-references: GB:AE005173; NID:gl1038497; PIDN:AAG27774.1; GSPDB:GN00141

C;Genetics:

A;Gene: T18T24.2

A;Map position: 1

Query Match 2.2%; Score 8; DB 2; Length 401;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241

Db 294 AGGGAGGG 301

RESULT 19

S01820

glycine-rich cell wall protein 1.8 precursor - kidney bean

C;Species: Phaseolus vulgaris (kidney bean)

C;Date: 30-Sep-1989 #sequence\_revision 19-May-1994 #text\_change 16-Jul-1999

C;Accession: S01820

R;Keller, B.; Sauer, N.; Lamb, C.J.

EMBO J. 7, 3625-3633, 1988

A;Title: Glycine-rich cell wall proteins in bean: gene structure and association of th

A;Reference number: S01820; MUID:89091109; PMID:3208742

A;Accession: S01820

A;Molecule type: DNA

A;Residues: 1-465 <KEL>

A;Cross-references: EMBL:X13596; NID:g21002; PIDN:CAA31932.1; PID:g21003

C;Comment: This protein is enriched in the cell wall fraction of young hypocotyls and

tyls.

C;Comment: Much of the sequence consists of tandemly repeated 22-residue segments with

C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

C;Keywords: cell wall; structural protein; tandem repeat

F;1-30/Domain: signal sequence #status predicted <SIG>

F;31-465/Product: glycine-rich cell wall protein 1.8 #status predicted <MAT>

Query Match 2.2%; Score 8; DB 1; Length 465;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241

Db 191 AGGGAGGG 198

C;Accession: T28910  
R;Martin, J.

submitted to the EMBL Data Library, December 1995  
A;Description: The sequence of *C. elegans* cosmid T26C11.

A;Reference number: Z20542

A;Accession: T28910

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-548 <MAR>

A;Cross-references: EMBL:U41017; PIDN:AAC48213.1; GSPDB:GN00028; CBSP:T26C11.4

A;Experimental source: strain Bristol N2; clone T26C11

C;Genetics:

A;Gene: CBSP:T26C11.4

A;Map position: X

A;Introns: 9/2; 85/1; 174/1; 202/1; 229/1; 261/1; 298/1; 337/1; 370/1; 393/1; 434/1; 471/1

Query Match 2.2%; Score 8; DB 2; Length 548;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 PQGASNAG 235

Db 195 PQGASNAG 202

Db

RESULT 21

G64874

probable membrane protein yciQ - *Escherichia coli* (strain K-12)

C;Species: *Escherichia coli*

C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C;Accession: G64874

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.

A;Title: The complete genome sequence of *Escherichia coli* K-12.

A;Reference number: A64720; MUID:97436617; PMID:9278503

A;Accession: G64874

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-631 <BLAT>

A;Cross-references: GB:AE000224; GB:U00096; NID:G1787509; PIDN:AAC74350.1; PID:G1787522;

A;Experimental source: strain K-12, substrain MGL655

C;Genetics:

A;Gene: yciQ

C;Superfamily: *Escherichia coli* probable membrane protein yciQ

C;Keywords: transmembrane protein

F;7-23/Domain: transmembrane #status predicted <TM1>

F;249-265/Domain: transmembrane #status predicted <TM2>

F;402-418/Domain: transmembrane #status predicted <TM3>

F;431-447/Domain: transmembrane #status predicted <TM4>

F;460-476/Domain: transmembrane #status predicted <TM5>

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 631;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241

Db 619 AGGAGGG 626

Db

RESULT 22

A45638

immunodominant microneme protein Etp100 - *Eimeria tenella*

C;Species: *Eimeria tenella*

C;Date: 22-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000

C;Accession: A45638

R;Tomley, F.M.; Clarke, L.E.; Kawazo, U.; Dijkema, R.; Kok, J.J.

Mol. Biochem. Parasitol. 49, 277-288, 1991

A;Title: Sequence of the gene encoding an immunodominant microneme protein of *Eimeria* te

A;Reference number: A45638; MUID:92131064; PMID:1775171

A;Accession: A45638

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-712 <TOM>

A;Cross-references: GB:AF032905; GB:M73495; NID:G2707732; PIDN:AAD03350.1; PID:G2707732;

A;Note: sequence extracted from NCBI backbone (NCBIN:77752; NCBIP:77756)

C;Superfamily: thrombospondin type 1 repeat homology; von Willebrand factor type A rei

F;18-218/Domain: von Willebrand factor type A repeat homology <VWA1>

F;238-296/Domain: thrombospondin type 1 repeat homology <THR1>

F;309-371/Domain: thrombospondin type 1 repeat homology <THR2>

F;372-432/Domain: thrombospondin type 1 repeat homology <THR3>

F;433-493/Domain: thrombospondin type 1 repeat homology <THR4>

F;494-556/Domain: thrombospondin type 1 repeat homology <THR5>

F;560-610/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 712;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NAGGGAGG 240

Db 612 NAGGGAGG 619

Db

RESULT 23

S55344

outer envelope membrane protein OBP75 precursor - garden pea

N;Alternate names: Chloroplast import-associated channel IAP75

C;Species: *Pisum sativum* (garden pea)

C;Date: 14-Oct-1995 #sequence\_revision 12-Apr-1996 #text\_change 21-Jul-2000

C;Accession: S55344; S55343; A55170; S51279

R;Tranel, P.J.; Froehlich, J.; Goyal, A.; Keegstra, K.

EMBO J. 14, 2436-2446, 1995

A;Title: A component of the chloroplastic protein import apparatus is targeted to the

A;Reference number: S55343; MUID:95300776; PMID:7781598

A;Accession: S55344

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-809 <TRA>

A;Cross-references: EMBL:X83767; NID:G633606; PIDN:CAAS8720.1; PID:G633607

A;Accession: S55343

A;Molecule type: protein

A;Residues: 136-150;185-196;360-367;532-545;642-660 <TRW>

R;Schnell, D.J.; Kessler, F.; Blobel, G.

Science 266, 1007-1012, 1994

A;Title: Isolation of components of the chloroplast protein import machinery.

A;Reference number: A55170; MUID:95063931; PMID:7973649

A;Accession: A55170

A;Molecule type: mRNA

A;Residues: 1-809 <SCH>

A;Cross-references: GB:L36858; NID:G576506; PIDN:AAA53275.1; PID:G576507

C;Genetics:

A;Genome: nuclear

C;Keywords: chloroplast; transmembrane protein

F;1-131/Domain: transit peptide (chloroplast) #status predicted <TNP>

F;132-809/Product: outer envelope membrane protein OBP75 #status predicted <MAT>

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 809;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241

Db 90 AGGAGGG 97

Db

RESULT 24

T02812

probable membrane protein L2759.4 [imported] - *Leishmania major* (strain Friedlin)

C;Species: *Leishmania major*

C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 19-May-2000

C;Accession: H81457; T02812

R;Myler, P.J.; Audleman, L.; deVos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.

Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999

A;Title: *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein

A;Reference number: AB1455; MUID:99178987; PMID:10077609  
A;Accession: H81457  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-823 <PYL>  
A;Cross-references: GB:AE001274; NID:g3264850; PIDN:AAC24635.1; PID:g2995588; GSPDB:GN00  
A;Experimental source: strain MHOM/IL/81/Friedlin  
C;Genetics:  
A;Gene: L2759.4  
A;Map position: 1

Query Match 2.2%; Score 8; DB 2; Length 823;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 ESTLEENK 63  
|||||  
Db 736 ESTLEENK 743

RESULT 25  
brkA prtein - Bordetella pertussis  
C;Species: Bordetella pertussis  
C;Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 08-Oct-1999  
C;Accession: I40329  
R;Fernandez, R.C.; Weiss, A.A.  
Infect. Immun. 62, 4727-4738, 1994  
A;Title: Cloning and sequencing of a Bordetella pertussis serum resistance locus.  
A;Reference number: I40328; MUID:95012680; PMID:7927748  
A;Accession: I40329  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1010 <RES>  
A;Cross-references: EMBL:U12276; NID:g562025; PIDN:AAAS1646.1; PID:g562026  
C;Genetics:  
A;Gene: brkA

Query Match 2.2%; Score 8; DB 2; Length 1010;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 QAPQPPVA 114  
|||||  
Db 43 QAPQPPVA 50

RESULT 26  
T13709  
diacylglycerol kinase (EC 2.7.1.107) - fruit fly (Drosophila melanogaster)  
C;Species: Drosophila melanogaster  
C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C;Accession: T13709  
R;Masai, I.; Okazaki, A.; Hosoya, T.; Hotta, Y.  
Proc. Natl. Acad. Sci. U.S.A. 90, 11157-11161, 1993  
A;Title: Drosophila retinal degeneration A gene encodes an eye-specific diacylglycerol k  
A;Reference number: Z17702; MUID:94068563; PMID:8248222  
A;Accession: T13709  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1454 <MAS>  
A;Cross-references: EMBL:D17315; NID:d1007485; PID:d1004654; PIDN:BA04135.1  
C;Genetics:  
A;Gene: rdga  
A;Cross-references: FlyBase:FBgn0003217  
C;Keywords: phosphotransferase

Query Match 2.2%; Score 8; DB 2; Length 1454;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
|||||

Db 764 AGGGAGGG 771

RESULT 27  
T14603  
hypotheical protein - Trypanosoma cruzi  
C;Species: Trypanosoma cruzi  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Sep-1999  
C;Accession: T14603; T14634  
R;Andersson, B.; Aslund, L.; Pettersson, U.  
submitted to the EMBL Data Library, March 1998  
A;Description: 93.4 kb of complete sequence from chromosome 3 of Trypanosoma cruzi.  
A;Reference number: Z18159  
A;Accession: T14603  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1718 <AND>  
A;Cross-references: EMBL:AF052832; NID:g3063540; PID:g3063541; PIDN:AAC14077.1  
A;Accession: T14634  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1718 <AN2>  
A;Cross-references: EMBL:AF052833; NID:g3063554; PID:g3063567; PIDN:AAC14102.1  
C;Genetics:  
A;Map position: 3

Query Match 2.2%; Score 8; DB 2; Length 1718;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
|||||  
Db 543 AGGGAGGG 550

RESULT 28  
D69426  
surface layer protein B (slgB-2) homolog - Archaeoglobus fulgidus  
C;Species: Archaeoglobus fulgidus  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000  
C;Accession: D69426  
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dode  
Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.  
Glock, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archa  
A;Reference number: A69250; MUID:98049343; PMID:9389475  
A;Accession: D69426  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-2425 <KLE>  
A;Cross-references: GB:AE001006; GB:AE000782; NID:g2689329; PIDN:AAB89834.1; PID:g264

Query Match 2.2%; Score 8; DB 2; Length 2425;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
|||||  
Db 1144 AGGGAGGG 1151

RESULT 29  
S77300  
hypotheical protein slr1403 - Synchocystis sp. (strain PCC 6803)  
C;Species: Synchocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C;Accession: S77300  
R;Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu

DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S77300  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-3016 <KAN>  
A;Cross-references: EMBL:D90907; GB:AB001339; NID:G1652618; PIDN:BAAL7634.1; PID:di01836  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 2.2%; Score 8; DB 2; Length 3016;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||  
Db 2629 AGGAGGG 2636

RESULT 30  
A71429  
hypothetical protein - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
A;Variety: Columbia  
C;Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998  
C;Accession: A71429  
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weiringer, T.; Pohl, T.M.; Terry, N.; Giel  
avanagh, T.; Hempel, S.; Kotter, F.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; And  
C.; Chalwatzie, N.  
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A;Reference number: A71400; MUID:98121113; PMID:9461215  
A;Accession: A71429  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-42 <BEV>  
A;Cross-references: GB:297340; NID:G2244950; PID:e327498; PID:G2244982  
C;Genetics:  
A;Map position: 4COP9-4G3845

Query Match 1.9%; Score 7; DB 2; Length 42;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
|||||  
Db 21 GGGAGGG 27

RESULT 31  
B49410  
t-complex polypeptide 1 homolog (peak 6a fraction) - rabbit (fragments)  
N;Alternate names: chaperonin homolog (peak 6a)  
C;Species: Oryctolagus cuniculus (domestic rabbit)  
C;Date: 21-Sep-1994 #sequence\_revision 18-Nov-1994 #text\_change 04-Sep-1998  
C;Accession: B49410  
R;Rommlaere, H.; Van Troys, M.; Gao, Y.; Melki, R.; Cowan, N.J.; Vandekerckhove, J.; Am  
Proc. Natl. Acad. Sci. U.S.A. 90, 11975-11979, 1993  
A;Title: Eukaryotic cytosolic chaperonin contains t-complex polypeptide 1 and seven rela  
A;Reference number: A49410; MUID:94089752; PMID:7903455  
A;Accession: B49410  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1-50 <ROM>  
A;Experimental source: reticulocyte  
A;Note: sequence modified after extraction from NCBI backbone  
A;Note: sequence extracted from NCBI backbone (NCBIP:141037)  
C;Superfamily: molecular chaperone t-complex-type

Query Match 1.9%; Score 7; DB 2; Length 50;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 MLVELSK 275  
|||||  
Db 31 MLVELSK 37

RESULT 32  
A39034  
neurogranin - bovine (fragment)  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 30-Aug-1991 #sequence\_revision 30-Aug-1991 #text\_change 24-Nov-1999  
C;Accession: A39034  
R;Baudier, J.; Deloulme, J.C.; Van Dorsselaer, A.; Black, D.; Matthes, H.W.D.  
J. Biol. Chem. 266, 229-237, 1991  
A;Title: Purification and characterization of a brain-specific protein kinase C subst:  
AP43) that corresponds to the protein kinase C phosphorylation site and the calmoduli  
A;Reference number: A39034; MUID:91093129; PMID:1824695  
A;Accession: A39034  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1-76 <BAU>  
C;Keywords: blocked amino end; phosphoprotein  
F;1/Modified site: blocked amino end (Met) (probably acetylated) #status experimental  
F;34/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 1.9%; Score 7; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGP 242  
|||||  
Db 67 GGAGGGP 73

RESULT 33  
A57288  
neurogranin - rat  
N;Alternate names: gene RC3 protein  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 01-Dec-1995 #sequence\_revision 01-Dec-1995 #text\_change 08-Oct-1999  
C;Accession: A57288; S36982; S36981  
R;Sato, T.; Xiao, D.M.; Li, H.; Huang, F.L.; Huang, K.P.  
J. Biol. Chem. 270, 10314-10322, 1995  
A;Title: Structure and regulation of the gene encoding the neuron-specific protein ki  
A;Reference number: A57288; MUID:95247744; PMID:7730337  
A;Accession: A57288  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-78 <SAT>  
A;Cross-references: GB:U22062; NID:924645; PIDN:AAA80223.1; PID:g924646  
R;Watson, J.B.; Battenberg, E.F.; Wong, K.K.; Bloom, F.E.; Sutcliffe, J.G.  
J. Neurosci. Res. 26, 397-408, 1990  
A;Title: Subtractive cDNA cloning of RC3, a rodent cortex-enriched mRNA encoding a no  
A;Reference number: S36982; MUID:91039409; PMID:2231781  
A;Accession: S36982  
A;Molecule type: mRNA  
A;Residues: 1-78 <WAT>  
A;Cross-references: EMBL:L09119; NID:G206594; PIDN:AAA42023.1; PID:G206595  
R;Huang, K.P.; Huang, F.L.; Chen, H.C.  
Arch. Biochem. Biophys. 305, 570-580, 1993  
A;Title: Characterization of a 7.5-kDa protein kinase C substrate (RC3 protein, neuro:  
A;Reference number: S36881; MUID:93384318; PMID:8080473  
A;Accession: S36881  
A;Molecule type: protein  
A;Residues: 2-78 <HUA>

Query Match 1.9%; Score 7; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGP 242  
Db 69 GGAGGGP 75

RESULT 34  
I47043  
neurogranin RC3 [imported] - goat  
C:Species: Capra sp.  
C:Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 20-Jun-2000  
C:Accession: I47043  
R:Piosik, P.A.; van Groenigen, M.; Ponne, N.J.; Bolhuis, P.A.; Baas, F.  
Brain Res. Mol. Brain Res. 29, 119-130, 1995  
A:Title: RC3/neurogranin structure and expression in the caprine brain in relation to c  
A:Reference number: I47043; MUID:95287801; PMID:7539519  
A:Accession: I47043  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-78 <PIO>  
A:Cross-references: GB:S78295; NID:g994839; PIDN:AAB34481.1; PID:g994841  
C:Genetics:  
A:Gene: RC3

Query Match 1.9%; Score 7; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGP 242  
Db 69 GGAGGGP 75

RESULT 35  
S49470  
hypothetical protein - Salmonella choleraesuis (fragment)  
C:Species: Salmonella choleraesuis  
C:Date: 20-Feb-1995 #sequence\_revision 26-Jul-1996 #text\_change 29-Sep-1999  
C:Accession: S49470  
R:Krause, M.; El-Gedaily, A.  
submitted to the EMBL Data Library, October 1994  
A:Reference number: S49470  
A:Accession: S49470  
A:Molecule type: DNA  
A:Residues: 1-94 <KRA>  
A:Cross-references: EMBL:X82129; NID:g558643; PIDN:CAA57639.1; PID:g558644  
A:Note: the source is designated as Salmonella enterica  
C:Superfamily: lipoprotein D

Query Match 1.9%; Score 7; DB 2; Length 94;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 GSTGTSS 87  
Db 65 GSTGTSS 71

RESULT 36  
F97355  
uncharacterized conserved protein, yuke/yfJA B. subtilis family CAC3713 [imported] - Clo  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: F97355  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: F97355  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-96 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK01633.1; PID:g15026818; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC3713

Query Match 1.9%; Score 7; DB 2; Length 96;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 ISVTPEE 325  
Db 4 ISVTPEE 10

RESULT 37  
S69752  
hypothetical protein YDR112w - Yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 23-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 29-Oct-1999  
C:Accession: S69752  
R:Murphy, L.; Shore, L.; Harris, D.  
submitted to the EMBL Data Library, March 1995  
A:Reference number: S52671  
A:Accession: S69752  
A:Molecule type: DNA  
A:Residues: 1-102 <MUR>  
A:Cross-references: EMBL:Z48750; GSPDB:GN00004; MIPS:YDR112w  
C:Genetics:  
A:Gene: MIPS:YDR112w  
A:Map position: 4R

Query Match 1.9%; Score 7; DB 2; Length 102;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 SKGKTSG 81  
Db 8 SKGKTSG 14

RESULT 38  
T45500  
hypothetical protein [imported] - Thiobacillus ferrooxidans (fragment)  
C:Species: Thiobacillus ferrooxidans  
C:Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T45500  
R:Oppon, J.C.; Sarnovsky, R.J.; Craig, N.L.; Rawlings, D.E.  
J. Bacteriol. 180, 3007-3012, 1998  
A:Title: A Tn7-like transposon is present in the glmUS region of the obligately chemo  
A:Reference number: Z22992; MUID:98269023; PMID:9603897  
A:Accession: T45500  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-102 <OPP>  
A:Cross-references: EMBL:AF032884; NID:g2653994; PIDN:AAC21668.1; PID:g2654001  
A:Experimental source: ATCC 33020

Query Match 1.9%; Score 7; DB 2; Length 102;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
Db 30 GGGAGGG 36

RESULT 39  
G86252  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 17-May-2002  
C:Accession: G86252  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.; Alons  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K

ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 C;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
 A;Reference number: A86141; MUID:21016719; PMID:11130712  
 A;Accession: G86252  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-108 <STO>  
 A;Cross-references: GB:AE005172; NID:g3157948; PIDN:AAC17631.1; GSPDB:GN00141  
 C;Genetics:  
 A;Map position: 1  
 C;Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology

Query Match 1.9%; Score 7; DB 2; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
 |||||  
 Db 97 GGGAGGG 103

RESULT 40  
 C55993  
 Hypothetical protein (pyp 3' region) - Ectothiorhodospira halophila (fragment)  
 C;Species: Ectothiorhodospira halophila  
 C;Date: 19-Jan-1996 #sequence\_revision 19-Jan-1996 #text\_change 08-Oct-1999  
 C;Accession: C55993  
 R;Baca, M.; Borgstahl, G.E.O.; Boissinot, M.; Burke, P.M.; Williams, D.R.; Slater, K.A.;  
 Biochemistry 33, 14369-14377, 1994  
 A;Title: Complete chemical structure of photoactive yellow protein: novel thioester-link  
 A;Reference number: A55993; MUID:95072006; PMID:7981196  
 A;Accession: C55993  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-112 <BAC>  
 A;Cross-references: GB:U17017; NID:G602427; PIDN:AAA61736.1; PID:G602430

Query Match 1.9%; Score 7; DB 2; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 TSGSTGT 85  
 |||||  
 Db 104 TSGSTGT 110

RESULT 41  
 S58445  
 Lipoprotein D - *Salmonella typhimurium* (fragment)  
 C;Species: *Salmonella typhimurium*  
 C;Date: 12-Feb-1998 #sequence\_revision 15-May-1998 #text\_change 29-Sep-1999  
 C;Accession: S58445  
 R;Robbe-Saule, V.; Coynault, C.; Norel, F.  
 FEWS Microbiol. Lett. 126, 171-176, 1995  
 A;Title: The live oral typhoid vaccine Ty21a is a rpoS mutant and is susceptible to vari  
 A;Reference number: S58445; MUID:95220644; PMID:7705608  
 A;Accession: S58445  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-119 <ROB>  
 A;Cross-references: EMBL:X81641; NID:G695757; PIDN:CAA57297.1; PID:G695758  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1994  
 C;Genetics:  
 A;Gene: nlpD  
 C;Superfamily: lipoprotein D

Query Match 1.9%; Score 7; DB 2; Length 119;

Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 GSGTSS 87  
 |||||  
 Db 90 GSGTSS 96

RESULT 42  
 DB3723  
 Hypothetical protein BH0588 [imported] - *Bacillus halodurans* (strain C-125)  
 C;Species: *Bacillus halodurans*  
 C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C;Accession: DB3723  
 R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* ar  
 A;Reference number: AB3650; MUID:20512582; PMID:11058132  
 A;Accession: DB3723  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-119 <STO>  
 A;Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04307.1; GSPDB:G  
 A;Experimental source: strain C-125  
 C;Genetics:  
 A;Gene: BH0588  
 C;Superfamily: *Escherichia coli* hypothetical 14K protein (cysG-trpS region)

Query Match 1.9%; Score 7; DB 2; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 ERAVEYL 183  
 |||||  
 Db 106 ERAVEYL 112

RESULT 43  
 A24272  
 Ig heavy chain precursor V region (RVH832) - rabbit (fragment)  
 C;Species: *Oryctolagus cuniculus* (domestic rabbit)  
 C;Date: 08-Aug-1987 #sequence\_revision 08-Aug-1987 #text\_change 21-Jan-2000  
 C;Accession: A24272  
 R;McCormack, W.T.; Laster, S.M.; Marzluff, W.F.; Roux, K.H.  
 Nucleic Acids Res. 13, 7041-7054, 1985  
 A;Title: Dynamic gene interactions in the evolution of rabbit Vh genes: a four codon  
 A;Reference number: A24272; MUID:86041895; PMID:2997735  
 A;Accession: A24272  
 A;Molecule type: DNA  
 A;Residues: 1-121 <MCC>  
 A;Cross-references: GB:X03054; NID:g1597; PIDN:CAA26862.1; PID:g1598  
 A;Note: this sequence was determined from the germline gene  
 C;Genetics:  
 A;Introns: 15/3  
 C;Superfamily: immunoglobulin V region; immunoglobulin homology  
 C;Keywords: heterotrimer; immunoglobulin  
 P:37-121/Domain: immunoglobulin homology <IMM>

Query Match 1.9%; Score 7; DB 2; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
 |||||  
 Db 26 GGGAGGG 32

RESULT 44  
 T36107  
 Probable serine/arginine rich protein - *Streptomyces coelicolor*  
 C;Species: *Streptomyces coelicolor*  
 C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C;Accession: T36107

R.Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, April 1999  
A;Reference number: Z21597  
A;Accession: T36107  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-172 <MUR>  
A;Cross-references: EMBL:AL049707; PIDN:CAB41273.1; GSPDB:GN000070; SCOEDB:SCE15.04  
A;Experimental source: EMBL: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SCE15.04

Query Match 1.9%; Score 7; DB 2; Length 172;

Best Local Similarity 100.0%; Pred. No. 50;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241

Db 51 GGGAGGG 57

RESULT 45

C82243

conserved hypothetical protein VC1075 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C;Species: Vibrio cholerae

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C;Accession: C82243

R;Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;

Chardonson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P.

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: C82243

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-180 <HEI>

A;Cross-references: GB:AR004189; GB:AE003852; NID:G9655541; PIDN:AAF94234.1; GSPDB:GN001

A;Experimental source: serogroup O1; strain N16961; biotype El Tor

C;Genetics:

A;Gene: VC1075

A;Map position: 1

C;Superfamily: conserved hypothetical protein ybhB

Query Match 1.9%; Score 7; DB 2; Length 180;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 PQGASNA 234

Db 151 PQGASNA 157

RESULT 46

I59203

gene HOX2.8 protein - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 15-Oct-1999

C;Accession: I59203

R;Lavellie, D.; Ducksworth, J.; Eves, E.; Gomes, G.; Keller, M.; Heller, P.; DeSimone, J.

Proc. Natl. Acad. Sci. U.S.A. 88, 7318-7322, 1991

A;Title: A homeodomain protein binds to gamma-globin gene regulatory sequences.

A;Reference number: I59203; MUID:913334457; PMID:1871139

A;Accession: I59203

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-182 <RES>

A;Cross-references: GB:M73999; NID:G184283; PIDN:AAB00778.1; PID:G184284

C;Superfamily: unassigned homeobox proteins; homeobox homology

C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

F;58-114/Domain: homeobox homology <HOX>

Query Match 1.9%; Score 7; DB 2; Length 182;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGG 240

Db 162 AGGAGG 168

RESULT 47

AF0830

probable membrane protein STY2838 [imported] - Salmonella enterica subsp. enterica ser

C;Species: Salmonella enterica subsp. enterica serovar Typhi

A;Note: This species has also been called Salmonella typhi

C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C;Accession: AF0830

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farra

l, S.; Mouton, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser

A;Reference number: AB0502; MUID:21534947; PMID:11677608

A;Accession: AF0830

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-195 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD02794.1; PID:G16503803; GSPDB:GN00176

C;Genetics:

A;Gene: STY2838

Query Match 1.9%; Score 7; DB 2; Length 195;

Best Local Similarity 100.0%; Pred. No. 57;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 GFLVWML 74

Db 49 GFLVWML 55

RESULT 48

T00799

hypothetical protein At2g32690 [imported] - Arabidopsis thaliana

N;Alternate names: hypothetical protein F24L7.17

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 16-Feb-2001

C;Accession: T00799; C84736

R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Ka

submitted to the EMBL Data Library, February 1998

A;Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.

A;Reference number: Z14204

A;Accession: T00799

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-201 <ROU>

A;Cross-references: EMBL:AC003974; NID:G2914688; PIDN:AAC04494.1; PID:G2914704

A;Experimental source: cultivar Columbia

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.B.; Umayam, L.; Tallon,

euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84736

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-201 <STO>

A;Cross-references: GB:AE002093; NID:G2914704; PIDN:AAC04494.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g32690; F24L7.17

A;Map position: 2

C;Superfamily: glycine-rich cell wall structural protein 1

```

Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GCGAGGG 241
    |||||
Db 138 GCGAGGG 144

RESULT 49
G01204
twist protein homolog - human
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 08-Oct-1999
C:Accession: G01204; JC6330
R:Cristofalo, V.J.
submitted to the EMBL Data Library, September 1995
A:Reference number: G06124
A:Accession: G01204
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-201 <CR1>
A:Cross-references: EMBL:X91662; NID:G999455; PIDN:CAA62850.1; PID:G999456
R:Wang, S.M.; Coljee, V.W.; Pignolo, R.J.; Rotenberg, M.O.; Cristofalo, V.J.; Sierra, F.
Gene 167, 83-92, 1997
A:Title: Cloning of the human twist gene: Its expression is retained in adult mesodermal
A:Reference number: JC6330; MUID:97225800; PMID:9073070
A:Accession: JC6330
A:Molecule type: mRNA
A:Residues: 1-52, 'A', 54-201 <WAN>
A:Cross-references: GB:X91662; NID:G999455
C:Comment: This protein is a basic helix-loop-helix DNA-binding transcription factor the
C:Genetics:
A:Gene: GDB:TWIST
A:Cross-references: GDB:6278991; OMIM:601622
A:Map position: 7p21-7p21
C:Keywords: DNA binding; transcription factor
F:108-120/Domain: DNA-binding #status Predicted <DNB>
F:121-201/Region: helix-loop-helix #status Predicted

Query Match 1.9%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GCGAGGG 241
    |||||
Db 83 GCGAGGG 89

RESULT 50
I53066
gene M-twist protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C:Accession: I53066; I66795
R:Wolf, C.; Thisee, C.; Scoetzel, C.; Thisee, B.; Gerlinger, P.; Perrin-Schmitt, F.
Dev. Biol. 143, 363-373, 1991
A:Title: The M-twist gene of Mus is expressed in subsets of mesodermal cells and is close
A:Reference number: I53066; MUID:91122450; PMID:1840517
A:Accession: I53066
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-206 <RES>
A:Cross-references: GB:M63649; NID:G202243; PIDN:AAA40514.1; PID:G202244
A:Accession: I66795
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-35, 'R', 37-90, 'P', 92-206 <RE2>
A:Cross-references: GB:M63650; NID:G202245; PIDN:AAA40515.1; PID:G202246
C:Genetics:
A:Gene: M-twist

Query Match 1.9%; Score 7; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 59;

```

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GCGAGGG 241
    |||||
Db 84 GCGAGGG 90

RESULT 51
T07854
germin-like protein (clone BnC4) - rape
C:Species: Brassica napus (rape)
C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C:Accession: T07854
R:Saenz-Vasquez, J.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z16177
A:Accession: T07854
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-207 <SAE>
A:Cross-references: EMBL:U21743; NID:G914910; PIDN:AAA86365.1; PID:G914911
A:Experimental source: cv. Samourai; 5 days old seedlings
C:Superfamily: germin

Query Match 1.9%; Score 7; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TVKTLKG 10
    |||||
Db 195 TVKTLKG 201

RESULT 52
H69037
conserved hypothetical protein MTH1284 - Methanobacterium thermoautotrophicum (strain
C:Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: H69037
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني,
K. S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fur
A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: H69037
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-209 <MTH>
A:Cross-references: GB:AE000893; GB:AE000666; NID:G2622375; PIDN:AAB85766.1; PID:G2622
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1284

Query Match 1.9%; Score 7; DB 2; Length 209;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 GPLDPLR 247
    |||||
Db 104 GPLDPLR 110

RESULT 53
CB7542
hypothetical protein CC2364 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: CB7542
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kol
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

```



A>Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: C87542  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-215 <STO>  
A:Cross-references: GB:AE005673; NID:gl3423895; PIDN:AAK24335.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC2364

Query Match 1.9%; Score 7; DB 2; Length 215;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 QPEGLPA 126  
|||||  
DB 124 QPEGLPA 130

## RESULT 54

T48482  
hypothetical protein T28J14.50 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jun-2000  
A:Accession: T48482  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: T24493  
A:Accession: T48482  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-216 <BEV>  
A:Cross-references: EMBL:AL163652  
A:Experimental source: cultivar Columbia; BAC clone T28J14  
C:Genetics:  
A:Map position: 5  
A:Note: T28J14.50  
C:Superfamily: Arabidopsis thaliana hypothetical protein T28J14.50

Query Match 1.9%; Score 7; DB 2; Length 216;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 EEQEAIG 330  
|||||  
DB 181 EEQEAIG 187

## RESULT 55

AF3273  
ribosomal protein L11 methyltransferase (EC 2.1.1.1-) [imported] - *Brucella melitensis* (e  
C:Species: *Brucella melitensis*  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
A:Accession: AF3273  
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A>Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AF3273  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-218 <KUR>  
A:Cross-references: GB:AE008917; PIDN:AAL51353.1; PID:gl7982052; GSPDB:GN00190  
A:Experimental source: strain 16M.  
C:Genetics:  
A:Gene: BMEI0171  
A:Map position: 1  
C:Keywords: methyltransferase

Query Match 1.9%; Score 7; DB 2; Length 218;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 318 AISVTPE 324  
|||||  
DB 128 AISVTPE 134

## RESULT 56

A53143  
testis-determining gene/SRY homolog - dunnart (*Sminthopsis macroura*) (fragment)  
C:Species: *Sminthopsis macroura*  
C:Date: 06-Oct-1994 #sequence\_revision 18-Nov-1994 #text\_change 24-Sep-1999  
A:Accession: A53143  
R:Poster, J.W.; Graves, J.A.  
Proc. Natl. Acad. Sci. U.S.A. 91, 1927-1931, 1994  
A>Title: An SRY-related sequence on the marsupial X chromosome: implications for the e  
A:Reference number: A53143; MUID:94173941; PMID:8127908  
A:Accession: A53143  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-224 <FOS>  
A:Cross-references: GB:S69429; NID:G545827; PIDN:AAB30154.1; PID:G545828  
A:Note: sequence inconsistent with nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:145890, NCBIP:145892)  
C:Superfamily: unassigned HMG box proteins; HMG box homology  
F:52-127/Domain: HMG box homology <HMG>

Query Match 1.9%; Score 7; DB 2; Length 224;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
|||||  
DB 41 GGGAGGG 47

## RESULT 57

A98166  
hypothetical protein AGR\_L\_563 [imported] - *Agrobacterium tumefaciens* (strain C58, Cer  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
A:Accession: A98166  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldma  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.  
Science 294, 2323-2328, 2001  
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium t*  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: A98166  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-228 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK88851.1; PID:gl5158615; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_563  
A:Map position: linear chromosome

Query Match 1.9%; Score 7; DB 2; Length 228;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 DNAASNL 139  
|||||  
DB 154 DNAASNL 160

## RESULT 58

AF3121  
hypothetical protein Atu4594 [imported] - *Agrobacterium tumefaciens* (strain C58, Dupon  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
A:Accession: AF3121  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCle  
; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AF3121  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-228 <RUR>  
 A:Cross-references: GB:AE008689; PIDN:AAL45388.1; PID:G17743085; GSPDB:GN00187  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu4594  
 A:Map position: linear chromosome

Query Match 1.9%; Score 7; DB 2; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 DNAASNL 139  
 |||||  
 Db 154 DNAASNL 160

## RESULT 59

T07358  
 ribosomal protein S3 - *Chlorella vulgaris* chloroplast  
 C:Species: chloroplast *Chlorella vulgaris*  
 C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 21-Jul-2000  
 C:Accession: T07358  
 R:Wakaugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, M.; Ito, S.; Tsudzuki, J.; Nakas Proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997  
 A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella*  
 A:Reference number: Z15985; MUID:97303241; PMID:9159184  
 A:Accession: T07358  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-231 <WAK>  
 A:Cross-references: EMBL:AB001684; NID:G2224352; PIDN:BAA58006.1; PID:G2224522  
 C:Genetics:  
 A:Gene: rps3  
 A:Genome: Chloroplast  
 C:Superfamily: Escherichia coli ribosomal protein S3  
 C:Keywords: chloroplast; ribosome

Query Match 1.9%; Score 7; DB 2; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 VQALRA 171  
 |||||  
 Db 154 VQALRA 160

## RESULT 60

G95989  
 hypothetical glycine-rich protein [imported] - *Sinorhizobium meliloti* (strain 1021) mega  
 C:Species: *Sinorhizobium meliloti*  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: G95989  
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N<sub>2</sub>-fixing endo  
 A:Reference number: A95842; MUID:21396508; PMID:11481431  
 A:Accession: G95989  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-234 <RUR>  
 A:Cross-references: GB:ALS1985; PIDN:CAC49583.1; PID:G15141070; GSPDB:GN00167  
 A:Experimental source: strain 1021, megaplasmid pSymb  
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau hebaul, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: SMB20889  
 A:Genome: plasmid

Query Match 1.9%; Score 7; DB 2; Length 234;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGG 240  
 |||||  
 Db 176 AGGAGG 182

## RESULT 61

I46082  
 CD8 alpha-chain - cat  
 C:Species: *Felis silvestris catus* (domestic cat)  
 C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 23-Jul-1999  
 C:Accession: I46082  
 R:Pecoraro, M.; Kawaguchi, Y.; Miyazawa, T.; Norimine, J.; Maeda, K.; Toyosaki, T.; T Immunology 81, 127-131, 1994  
 A:Title: Isolation, sequence and expression of a cDNA encoding the alpha-chain of the  
 A:Reference number: I46082; MUID:94178799; PMID:8132208  
 A:Accession: I46082  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-239 <PEC>  
 A:Cross-references: GB:D16536; NID:G485385; PIDN:BAA03973.1; PID:G485386  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology

Query Match 1.9%; Score 7; DB 2; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 APITTSQ 120  
 |||||  
 Db 153 APITTSQ 159

## RESULT 62

T46968  
 flavoprotein [imported] - *Paracoccus denitrificans* (fragment)  
 C:Species: *Paracoccus denitrificans*  
 C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 21-Jul-2000  
 C:Accession: T46968  
 R:Wodara, C.; Bardischewsky, F.; Friedrich, C.G.  
 J. Bacteriol. 179, 5014-5023, 1997  
 A:Title: Cloning and characterization of sulfite dehydrogenase, two c-type cytochromes  
 A:Reference number: Z24324; MUID:97405897; PMID:9260941  
 A:Accession: T46968  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-247 <WOD>  
 A:Cross-references: EMBL:X79242; NID:G2253074; PIDN:CAA55826.1; PID:G2222781  
 A:Experimental source: strain GB17  
 C:Genetics:  
 A:Gene: soxP

Query Match 1.9%; Score 7; DB 2; Length 247;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGG 241  
 |||||  
 Db 36 GGGAGG 42

## RESULT 63

T33230  
 hypothetical protein T10B5.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T33230  
 R:Giesel, C.; Bradshaw, H.; O'Brien, D.  
 A:Description: The sequence of C. elegans cosmid T10B5.  
 A:Reference number: 221305  
 A:Accession: T33230  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-248 <GEI>  
 A:Cross-references: EMBL:AF067947; PIDN:AA19227.1; GSPDB:GN00023; CESP:T10B5.3  
 A:Experimental source: strain Bristol N2; clone T10B5  
 C:Genetics:  
 A:Gene: CESP:T10B5.3  
 A:Map position: 5  
 A:Introns: 65/2; 141/2

Query Match 1.9%; Score 7; DB 2; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 76 KGKTSGS 82

Db 72 KGKTSGS 78

## RESULT 64

B33144  
 homeotic protein Ultrabithorax - house fly (fragment)  
 C:Species: Musca domestica (house fly)  
 C>Date: 30-Nov-1990 #sequence\_revision 20-Sep-1991 #text\_change 18-Jun-1993  
 C:Accession: B33144  
 R:Wilde, C.D.; Akam, M.  
 EMBL J. 6, 1393-1401, 1987  
 A:Title: Conserved sequence elements in the 5' region of the Ultrabithorax transcription  
 A:Reference number: A33144  
 A:Accession: B33144  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-249 <WIL>  
 A:Cross-references: GB:X05178; GB:Y00066

Query Match 1.9%; Score 7; DB 2; Length 249;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241

Db 108 GGGAGGG 114

## RESULT 65

B84777  
 hypothetical protein At2g36120 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: B84777  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;  
 euss, D.; Nierman, W.C.; White, C.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: B84777  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-255 <STO>  
 A:Cross-references: GB:AE002093; NID:G4678224; PIDN:AAD26969.1; GSPDB:GN00139

## C:Genetics:

A:Gene: At2g36120  
 A:Map position: 2  
 C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homolog

Query Match 1.9%; Score 7; DB 2; Length 255;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241

Db 89 GGGAGGG 95

## RESULT 66

CIRBL  
 calpain (EC 3.4.22.17) small chain - rabbit  
 A:Alternate names: calcium-activated neutral proteinase (CANP); calpain light chain; c  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C>Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999  
 C:Accession: A24816  
 R:Emori, Y.; Kawasaki, H.; Imajoh, S.; Kawashima, S.; Suzuki, K.  
 J. Biol. Chem. 261, 9472-9476, 1986  
 A:Title: Isolation and sequence analysis of cDNA clones for the small subunit of rabbi  
 A:Reference number: A24816; MUID:86250903; PMID:3013892  
 A:Accession: A24816  
 A:Molecule type: mRNA  
 A:Residues: 1-266 <EMO>  
 A:Cross-references: GB:M13364; NID:G164875; PIDN:AA81565.1; PID:G164876  
 C:Complex: heterodimer of L (large) and S (small) chains  
 C:Function:

A:Description: catalyzes the hydrolysis of peptides  
 A:Note: Cleaves preferentially after tyrosine, methionine, or arginine residues and be  
 C:Superfamily: calpain small chain; calmodulin repeat homology  
 C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer; h  
 F:1-54/Domain: glycine-rich <GLY>  
 F:94-125/Domain: calmodulin repeat homology <EF1>  
 F:137-169/Domain: calmodulin repeat homology <EF2>  
 F:170-199/Domain: calmodulin repeat homology <EF3>  
 F:202-234/Domain: calmodulin repeat homology <EF4>  
 F:235-266/Domain: calmodulin repeat homology <EF5>

Query Match 1.9%; Score 7; DB 1; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241

Db 47 GGGAGGG 53

## RESULT 67

S09860  
 hypothetical protein UL112 - human cytomegalovirus (strain AD169) (fragment)  
 C:Species: human cytomegalovirus, human herpesvirus 5  
 A:Note: host Homo sapiens (man)  
 C>Date: 07-Sep-1990 #sequence\_revision 31-Dec-1991 #text\_change 27-Jan-1995  
 C:Accession: S09860  
 R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohnl, R.; Brown, C.M.; Cerny, R.; Horsnell, T.  
 M.; Barrett, B.G.  
 Curr. Top. Microbiol. Immunol. 154, 125-169, 1990  
 A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovir  
 A:Reference number: S09749; MUID:90269039; PMID:2161319  
 A:Accession: S09860  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-268 <CHE>  
 A:Cross-references: EMBL:X17403  
 A:Note: this sequence was submitted to the EMBL Data Library, December 1989

Query Match 1.9%; Score 7; DB 2; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
 |||||  
 Db 206 GGGAGGG 212

RESULT 68  
 G83023  
 3',5'-cyclic-nucleotide phosphodiesterase (EC 3.1.4.17) cpdA-type PA4969 [similarity] -  
 N;Alternate names: icc protein  
 C;Species: Pseudomonas aeruginosa  
 C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 24-Aug-2001  
 C;Accession: G83023  
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A;Reference number: A82950; MUID:20437337; PMID:10984043  
 A;Accession: G83023  
 A;Molecule type: DNA  
 A;Residues: 1-272 <STO>  
 A;Cross-references: GB:AE004910; GB:AE004091; NID:g9951251; PIDN:AA08354.1; GSPDB:GN001  
 A;Experimental source: strain PA01  
 C;Genetics:  
 A;Gene: PA4969  
 C;Superfamily: 3',5'-cyclic-nucleotide phosphodiesterase cpdA; 3',5'-cyclic-nucleotide p  
 C;Keywords: iron; metalloprotein; phosphoric diester hydrolase  
 F;15-202/Domain: 3',5'-cyclic-nucleotide phosphodiesterase cpdA homology <CPDA>  
 F;15-95/Domain: phosphoesterase core homology <PEC>

Query Match 1.9%; Score 7; DB 1; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 HSNTPAT 96  
 |||||  
 Db 4 HSNTPAT 10

RESULT 69  
 A83583  
 probable biotin synthesis protein BioC PA0503 [imported] - Pseudomonas aeruginosa (strain  
 C;Species: Pseudomonas aeruginosa  
 C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C;Accession: A83583  
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A;Reference number: A82950; MUID:20437337; PMID:10984043  
 A;Accession: A83583  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-274 <STO>  
 A;Cross-references: GB:AE004487; GB:AE004091; NID:g9946361; PIDN:AA03892.1; GSPDB:GN001  
 A;Experimental source: strain PA01  
 C;Genetics:  
 A;Gene: PA0503

Query Match 1.9%; Score 7; DB 2; Length 274;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 QPEGLPA 126  
 |||||  
 Db 255 QPEGLPA 261

RESULT 70  
 AH2885  
 hypothetical protein dppD [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C;Species: Agrobacterium tumefaciens  
 C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C;Accession: AH2885  
 R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCle  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
 ster, E.W.  
 A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A;Reference number: A82577; MUID:21608550; PMID:11743193  
 A;Accession: AH2885  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-276 <KUR>  
 A;Cross-references: GB:AE008688; PIDN:AA43502.1; PID:g17741010; GSPDB:GN00186  
 A;Experimental source: strain C58 (Dupont)  
 C;Genetics:  
 A;Gene: dppD  
 A;Map position: circular chromosome

Query Match 1.9%; Score 7; DB 2; Length 276;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 ILRLIEE 286  
 |||||  
 Db 190 ILRLIEE 196

RESULT 71  
 E97661  
 dipeptide transporter dppD homolog (AB026907) [imported] - Agrobacterium tumefaciens  
 C;Species: Agrobacterium tumefaciens  
 C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C;Accession: E97661  
 R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm  
 A.; Liu, F.; Mollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, I  
 Science 294, 2323-2328, 2001  
 A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
 A;Reference number: A97359; MUID:21608551; PMID:11743194  
 A;Accession: E97661  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-276 <KUR>  
 A;Cross-references: GB:AE007869; PIDN:AAK88246.1; PID:g15157704; GSPDB:GN00169  
 C;Genetics:  
 A;Gene: AGR\_C\_4569  
 A;Map position: circular chromosome

Query Match 1.9%; Score 7; DB 2; Length 276;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 ILRLIEE 286  
 |||||  
 Db 190 ILRLIEE 196

RESULT 72  
 S74256  
 homeotic protein six3-alpha - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 02-Jul-1998  
 C;Accession: S74256  
 R;Kawakami, K.; Ohto, H.; Takizawa, T.; Saito, T.  
 FEBS Lett. 393, 259-263, 1996  
 A;Title: Identification and expression of six family genes in mouse retina.  
 A;Reference number: S74253; MUID:96409319; PMID:8814301  
 A;Accession: S74256  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-284 <KAW>

A:Cross-references: EMBL:D83147  
A:Experimental source: retina; strain BALB/c  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regulation  
F:208-264/Domain: homeobox homology <HGX>

Query Match 1.9%; Score 7; DB 2; Length 284;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241

Db 42 GGGAGGG 48

#### RESULT 73

T23416

hypothetical protein K07F5.11 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T23416

R:Hembry, C.

submitted to the EMBL Data Library, March 1996

A:Reference number: Z19738

A:Accession: T23416

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-290 <WIL>

A:Cross-references: EMBL:Z70284; PIDN:CAA94280.1; GSPDB:GN00022; CESP:K07F5.11

A:Experimental source: clone K07F5

C:Genetics:

A:Gene: CESP:K07F5.11

A:Map position: 4

A:Introns: 89/1; 257/1

C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 1.9%; Score 7; DB 2; Length 290;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241

Db 275 GGGAGGG 281

#### RESULT 74

S31415

glycine-rich protein GRP22 - rape.

C:Species: Brassica napus (rape)

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S31415

R:Bergeron, D.; Boivin, R.; Raszczynski, C.L.; Bellemare, G.

submitted to the EMBL Data Library, August 1992

A:Description: Characterization and expression of a gene family encoding glycine-rich pr

A:Reference number: S31415

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-231 <BER>

A:Cross-references: EMBL:Z15045; NID:G17820; PIDN:CAA78762.1; PID:G17821

C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 1.9%; Score 7; DB 1; Length 291;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241

Db 85 GGGAGGG 91

#### RESULT 75

T00503

probable MYB family transcription factor [imported] - Arabidopsis thaliana

N:Alternate names: myb-related protein homolog T20D16.8

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 23-Mar-2001

C:Accession: T00503; H84622

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crooby, M.L.; Brandon, R.C.; Sykes, S.M.; Ke

submitted to the EMBL Data Library, November 1997

A:Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.

A:Reference number: Z14164

A:Accession: T00503

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-309 <ROU>

A:Cross-references: EMBL:AC002391; NID:G2642427; PIDN:AAB87103.1; PID:G2642435

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: H84622

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-309 <STO>

A:Cross-references: GB:AE002093; NID:G2642435; PIDN:AAB87103.1; GSPDB:GN00139

C:Genetics:

A:Map position: 2

C:Superfamily: Arabidopsis myb-related 33.2K protein; myb DNA-binding repeat homology

F:8-59/Domain: myb DNA-binding repeat homology <MYB>

F:60-110/Domain: myb DNA-binding repeat homology <MYB2>

Query Match 1.9%; Score 7; DB 2; Length 309;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241

Db 277 GGGAGGG 283

Search completed: December 17, 2003, 06:36:01

Job time : 21.6624 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:21:26 ; Search time 10.4735 Seconds  
(without alignments)  
1652.347 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 368

Sequence: 1 MKLTVKTLKGTGHEIRVQPN.....CDRNEELAANYLLEHAGEED 368

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.4	368	1 RH23 SCHPO	O74803 schizosacch
2	8	2.2	331	1 SHX2 HUMAN	O60902 homo sapien
3	8	2.2	331	1 SHX2 MOUSE	P70390 mus musculus
4	8	2.2	349	1 GRP ARATH	P27483 arabidopsis
5	8	2.2	384	1 GRP PETHY	P09789 petunia hyb
6	8	2.2	456	1 FXD1 MOUSE	O61345 mus musculus
7	8	2.2	465	1 GRP2 PHAVU	P10496 phaseolus v
8	8	2.2	497	1 IRBP HUMAN	O60548 homo sapien
9	8	2.2	610	1 YC1Q ECOLI	P49194 mus musculus
10	8	2.2	631	1 YC1Q ECOLI	P45848 escherichia
11	8	2.2	828	1 CADM HUMAN	O90199 homo sapien
12	8	2.2	1454	1 KDGE DROME	O09103 drosophila
13	7	1.9	76	1 NEUG BOVIN	P35722 bos taurus
14	7	1.9	78	1 NEUG CAPHI	P54877 capra hircu
15	7	1.9	78	1 NEUG HUMAN	O92686 homo sapien
16	7	1.9	78	1 NEUG RAT	O04940 rattus norv
17	7	1.9	96	1 Y1B3 CLOAB	P34159 clostridium
18	7	1.9	97	1 NLDP YEREN	P47764 yersinia en
19	7	1.9	110	1 PER DROER	O26288 drosophila
20	7	1.9	141	1 HBB DASAK	P56592 dasyatis ak
21	7	1.9	202	1 TWS1 HUMAN	O15672 homo sapien
22	7	1.9	206	1 TWS1 MOUSE	P26687 mus musculus
23	7	1.9	207	1 GLP1 BRANA	P46271 brassica na
24	7	1.9	231	1 RRS3 CHLVU	P56365 chlorella v
25	7	1.9	239	1 CDBA FELCA	P41688 felis silve
26	7	1.9	248	1 HMXK MUSDO	O25451 musca domes
27	7	1.9	264	1 GSHI HUMAN	O90482 homo sapien
28	7	1.9	266	1 CANS RABIT	P06813 oryctolagus
29	7	1.9	268	1 EP34 HCMVA	P16768 human cytom
30	7	1.9	303	1 CDK4 FIG	P79432 sus scrofa
31	7	1.9	306	1 RALY HUMAN	O90km9 homo sapien
32	7	1.9	320	1 DDL XYLFA	O9pf79 xylella faa
33	7	1.9	323	1 HXDB MOUSE	P23013 mus musculus

# ALIGNMENTS

RESULT 1

34	7	1.9	329	1 A32C DROME	O46203 drosophila
35	7	1.9	329	1 TAL MOUSE	P22091 mus musculus
36	7	1.9	332	1 SIX3 HUMAN	O95343 homo sapien
37	7	1.9	333	1 SIX3 MOUSE	O62333 mus musculus
38	7	1.9	339	1 SYPA CLOAB	Q97gk9 clostridium
39	7	1.9	342	1 ROAL SCHAM	P21522 schistocercu
40	7	1.9	367	1 BET3 MESAU	O09029 mesocricetu
41	7	1.9	373	1 NLDP SALTI	O56131 salmoneilla
42	7	1.9	375	1 GLK1 TRIVA	O9gtw9 trichomonas
43	7	1.9	376	1 DNAB LISIN	O92bn9 listeria in
44	7	1.9	377	1 DNAB LISMO	O985a3 listeria mo
45	7	1.9	377	1 NLDP SALDU	P39700 salmonella
46	7	1.9	377	1 NLDP SALTY	P40827 salmonella
47	7	1.9	377	1 RRG1 HUMAN	O06587 homo sapien
48	7	1.9	378	1 DNJ2 STRCO	P33648 escherichia
49	7	1.9	379	1 NLDP ECOLI	O9rdd4 streptomyce
50	7	1.9	383	1 TCF7 HUMAN	P36402 homo sapien
51	7	1.9	385	1 RO32 XENLA	P51992 xenopus lae
52	7	1.9	389	1 DNAB HALCU	O34135 halobacteri
53	7	1.9	391	1 DNAB HALN1	O9hry3 halobacteri
54	7	1.9	391	1 PCL ECTHA	P42516 ectothiorho
55	7	1.9	393	1 HXA8 HUMAN	P31260 homo sapien
56	7	1.9	406	1 HEM1 AQUAE	O67314 aquifex aeo
57	7	1.9	413	1 YB44 MYCPN	P75142 mycoplasma
58	7	1.9	416	1 SO DROME	Q27350 drosophila
59	7	1.9	417	1 PCL RHOC	O69140 rhodobacter
60	7	1.9	421	1 CCG8 RAT	O8vhw5 rattus norv
61	7	1.9	423	1 CCG8 MOUSE	Q8vhw2 mus musculus
62	7	1.9	423	1 MKR2 SERQU	Q9dd48 seriola qui
63	7	1.9	425	1 CCG8 HUMAN	Q8wx85 homo sapien
64	7	1.9	429	1 SR54 THEAQ	O07347 thermus aqu
65	7	1.9	433	1 ELT2 CABEL	Q10655 caenorhabdi
66	7	1.9	437	1 KLCM MOUSE	P08730 mus musculus
67	7	1.9	440	1 FXGA CHICK	O98937 gallus gall
68	7	1.9	440	1 SOX4 MOUSE	Q08831 mus musculus
69	7	1.9	441	1 SX11 HUMAN	P35716 homo sapien
70	7	1.9	443	1 COBB METJA	Q58816 methanococc
71	7	1.9	454	1 GLNA HALVO	P43386 halobacteri
72	7	1.9	465	1 FXD1 HUMAN	O16676 homo sapien
73	7	1.9	465	1 FXD3 MOUSE	Q61060 mus musculus
74	7	1.9	467	1 SYC THETN	Q8r7t3 thermoanaer
75	7	1.9	486	1 KIC0 XENLA	P02537 xenopus lae
76	7	1.9	491	1 AMPG ECOLI	P36670 escherichia
77	7	1.9	495	1 P033 MOUSE	P31361 mus musculus
78	7	1.9	497	1 P033 RAT	Q63262 rattus norv
79	7	1.9	500	1 P033 HUMAN	P20264 homo sapien
80	7	1.9	503	1 ATPA ODOSI	Q00820 odontella s
81	7	1.9	510	1 ATPA RICCN	Q92986 rickettsia
82	7	1.9	512	1 YF00 MYCPN	O50288 rickettsia
83	7	1.9	527	1 YF00 MYCPN	P75287 mycoplasma
84	7	1.9	528	1 ATPA STRLI	P50001 streptomyce
85	7	1.9	532	1 ZIC2 HUMAN	O95409 homo sapien
86	7	1.9	536	1 TCFD FUGRU	P53451 fugu rubrip
87	7	1.9	539	1 TCFD HUMAN	P50991 homo sapien
88	7	1.9	539	1 TCFD MOUSE	P80315 mus musculus
89	7	1.9	540	1 TCFD CABEL	P47208 caenorhabdi
90	7	1.9	549	1 DSX DROME	P23023 drosophila
91	7	1.9	556	1 SR54 CANAL	O42816 candida alb
92	7	1.9	557	1 PKSJ BACSU	P40806 bacillus su
93	7	1.9	572	1 ACSA BACSU	P39062 bacillus su
94	7	1.9	576	1 P80C HUMAN	P38432 homo sapien
95	7	1.9	590	1 K2C5 HUMAN	P13647 homo sapien
96	7	1.9	612	1 ADF1 CANAL	P46589 candida alb
97	7	1.9	622	1 KICI HUMAN	P35527 homo sapien
98	7	1.9	624	1 STS MOUSE	P50427 mus musculus
99	7	1.9	627	1 TES1 MOUSE	O70146 mus musculus
100	7	1.9	628	1 TES1 RAT	Q63572 rattus norv

```

RH23_SCHPO STANDARD; PRT; 368 AA.
ID RH23_SCHPO
AC 074803;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UV excision repair protein rhp23 (RAD23 homolog).
GN RHP23 OR SPBC2D10.12.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC STRAIN=SP223;
RC STRAIN=21648918; PubMed=11788722;
RA Elder R.T., Song X.-Q., Chen M., Hopkins K.M., Lieberman H.B.,
RA Zhao Y.;
RT "Involvement of rhp23, a Schizosaccharomyces pombe homolog of the
RT human HHR23A and Saccharomyces cerevisiae RAD23 nucleotide excision
RT repair genes, in cell cycle control and protein ubiquitination.";
RL Nucleic Acids Res. 30:581-591(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Spours J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=20119127; PubMed=10652237;
RA Lombaerts M., Goeloe J.I., den Dulk H., Brandema J.A., Brouwer J.;
RT "Identification and characterization of the rhp23(+) DNA repair gene
RT in Schizosaccharomyces pombe.";
RL Biochem. Biophys. Res. Commun. 268:210-215(2000).
CC -1- FUNCTION: Involved in postreplication repair of UV-damaged DNA.
CC strand on replication functions in gap-filling of a daughter
CC strand on replication of damaged DNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 1 ubiquitin-like domain.
CC -1- SIMILARITY: Contains 2 UBA domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/

```

```

or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF174293; AAD51975.1; -
CC DR EMBL; AL031788; CAA21170.1; -
CC DR PIR; T40115; T40115.
CC DR HSSP; P54725; 1DV0.
CC DR GeneDB_Sfombe; SPBC2D10.12; -.
CC DR InterPro; IPR004806; Rad23.
CC DR InterPro; IPR006636; STIL.
CC DR InterPro; IPR000449; UBA domain.
CC DR InterPro; IPR000626; Ubiquitin.
CC DR Pfam; PF00627; UBA; 2.
CC DR Pfam; PF00240; ubiquitin; 1.
CC DR SMART; SM00727; STIL; 1.
CC DR SMART; SM00165; UBA; 2.
CC DR SMART; SM00213; UBQ; 1.
CC DR TIGRFAMs; TIGR00601; rad23; 1.
CC DR PROSITE; PS50033; UBIQUITIN 2; 1.
CC DR PROSITE; PS50033; UBIQUITIN 2; 1.
KW DNA damage; DNA repair; Nuclear protein; Repeat.
FT DOMAIN 1 77 UBIQUITIN-LIKE.
FT DOMAIN 135 185 UBA 1.
FT DOMAIN 320 360 UBA 2.
FT DOMAIN 119 122 POLY-ALA.
FT DOMAIN 205 208 POLY-GLN.
SQ SEQUENCE 368 AA; 40135 MW; 5CE75EB7E190EFD4 CRC64;

Query Match 2.4%; Score 9; DB 1; Length 368;
Best Local Similarity 100.0%; Pred. No. 0.42; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 352 NEELAANYL 360
DB 349 NEELAANYL 357
|||||||
|||||||

RESULT 2
SHX2_HUMAN STANDARD; PRT; 331 AA.
ID SHX2_HUMAN STANDARD; PRT; 331 AA.
AC O60902; O60465; O60467; O60903;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Short stature homeobox protein 2 (Paired-related homeobox protein
DE SHOT) (Homeobox protein Ogl2X).
GN SHOX2 OR SHOT OR OGI2X.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Muscle;
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish N.K.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uadiri T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

```

RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3].  
RP SEQUENCE OF 116-331 FROM N.A. (ISOFORM 2).  
RC TISSUE=Craniiofacial;  
RX MEDLINE=98133920; PubMed=9466998;  
RA Semina E.V., Reiter R.S., Murray J.C.;  
RT "A new human homeobox gene OGL2X is a member of the most conserved  
RT homeobox gene family and is expressed during heart development in  
RT mouse.";  
RL Hum. Mol. Genet. 7:415-422(1998).  
CC -!- FUNCTION: May be a growth regulator and have a role in specifying  
CC neural systems involved in processing somatosensory information,  
CC as well as in face and body structure formation.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1; Synonyms=SHOX2A, SHOTA;  
CC IsoId=O60902-1; Sequence=Displayed;  
CC Name=2; Synonyms=SHOX2B, SHOTB, OGL2XB;  
CC IsoId=O60902-2; Sequence=VSP\_002288;  
CC -!- TISSUE SPECIFICITY: Expressed in heart, skeletal muscle, liver,  
CC lung, bone marrow fibroblast, pancreas and placenta.  
CC -!- DEVELOPMENTAL STAGE: Expressed during craniofacial development as  
CC well as in heart.  
CC -!- SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY. BICOID  
CC SUBFAMILY.  
CC -!- SIMILARITY: Contains 1 homeobox domain.  
CC -!- SIMILARITY: Contains 1 OAR domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
CC -----  
DR EMBL; AJ002367; CAA05341.1; ALT\_INIT.  
DR EMBL; AJ002368; CAA05342.1; ALT\_INIT.  
DR EMBL; BC008829; AAH08829.1; -.  
DR EMBL; AF022654; AAC39662.1; ALT\_INIT.  
DR EMBL; AF023203; AAC39663.1; -.  
DR HSP; P06601; IRLJ.  
DR TRANSFAC; T04223; -.  
DR TRANSFAC; T04224; -.  
DR Genew; HGNC:10854; SHOX2.  
DR MIM; 602504; -.  
DR GO; GO:0007507; P:heart development; TAS.  
DR GO; GO:0007399; P:neurogenesis; TAS.  
DR GO; GO:0001501; P:skeletal development; TAS.  
DR InterPro; IPR003654; Homeo OAR.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR000047; HTH lambrpressor.  
DR Pfam; PF00046; homeobox; 1.  
DR Pfam; PF03826; OAR; 1.  
DR PRINTS; PR00031; HTHREPRESSR.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEOBOX 1; 1.  
DR PROSITE; PS00071; HOMEOBOX 2; 1.  
DR PROSITE; PS00803; OAR; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Alternative splicing.

FT DNA\_BIND 140 199 HOMEOBOX.  
FT DOMAIN 313 326 OAR.  
FT DOMAIN 60 86 POLY-GLY.  
FT VARSPPLIC 235 246 Missing (in isoform 2).  
FT CONFLICT 115 115 /FTId=VSP\_002288.  
FT CONFLICT 125 125 E -> EGRRPPTKAEVQATLLPGEAFREL (IN REF.  
FT CONFLICT 244 244 2).  
FT CONFLICT 312 312 E -> D (IN REF. 1; CAA05341).  
FT CONFLICT 325 325 P -> S (IN REF. 2).  
FT CONFLICT 325 325 D -> N (IN REF. 2 AND 3).  
FT CONFLICT 325 325 H -> L (IN REF. 3; AAC39663).  
SQ SEQUENCE 331 AA, 34964 MW, 55431B073B3B2250 CRC64;  
Query Match 2.2%; Score 8; DB 1; Length 331;  
Best Local Similarity 100.0%; Pred. No. 4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 234 AGGAGGG 241  
Db 83 AGGAGGG 90  
|||||  
|||  
RESULT 3  
SHX2\_MOUSE STANDARD; PRT; 331 AA.  
ID SHX2\_MOUSE  
AC P70390; P70369;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Short stature homeobox protein 2 (Homeobox protein Ogl2X) (OG-12)  
DE (Paired family homeodomain protein Prx3).  
GN SHOX2 OR OGL2X OR PRX3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX TISSUE=Embryo;  
RX MEDLINE=98058757; PubMed=9371788;  
RA van Schaick H.S.A., Smidt M.P., Roverscalli A.C., Luijten M.,  
RA van der Kleij A.A.M., Asoh S., Kozak C.A., Nirenberg M.W.,  
RA Burbach J.P.H.;  
RT "Homeobox gene Prx3 expression in rodent brain and extraneural  
RT tissues.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:12993-12998(1997).  
RN [2]  
RP SEQUENCE OF 116-331 AND 235-331 FROM N.A. (ISOFORMS 1 AND 2).  
RC STRAIN=BALB/c; TISSUE=Embryo, and Liver;  
RX MEDLINE=97008065; PubMed=8855241;  
RA Roverscalli A.C., Asoh S., Nirenberg M.W.;  
RT "Cloning and characterization of four murine homeobox genes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:10691-10696(1996).  
RN [3]  
RP DEVELOPMENTAL EXPRESSION.  
RX MEDLINE=98133920; PubMed=9466998;  
RA Semina E.V., Reiter R.S., Murray J.C.;  
RT "A new human homeobox gene OGL2X is a member of the most conserved  
RT homeobox gene family and is expressed during heart development in  
RT mouse.";  
RL Hum. Mol. Genet. 7:415-422(1998).  
CC -!- FUNCTION: May be a growth regulator and have a role in specifying  
CC neural systems involved in processing somatosensory information,  
CC as well as in face and body structure formation. May also have a  
CC role in heart development.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1; Synonyms=OGL12A, PRX3A;  
CC IsoId=P70390-1; Sequence=Displayed;  
CC Name=2; Synonyms=OGL12B, PRX3B;  
CC IsoId=P70390-2; Sequence=VSP\_002289;  
CC -!- TISSUE SPECIFICITY: Highly expressed in striated muscle followed



CC by liver, kidney, testis, brain, heart, lung and spleen.  
 CC -!- DEVELOPMENTAL STAGE: Expressed from E9 TO E16 day in the heart,  
 CC otic region, maxillary and mandibular components of the first  
 CC branchial arch, naeal processes, eyelid, midbrain, medulla  
 CC oblongata, limbs, dorsal root ganglia and genital tubercle. Also  
 CC expressed in non-neuronal structures around the oral cavity and in  
 CC hip and shoulder regions and in mesenchyme surrounding the  
 CC vertebrate.  
 CC -!- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY. BICOID  
 CC SUBFAMILY.  
 CC -!- SIMILARITY: Contains 1 homeobox domain.  
 CC -!- SIMILARITY: Contains 1 OAR domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U66918; AAC52833.1; -;  
 CC EMBL; U67055; AAC52834.1; -;  
 CC EMBL; U65071; -; NOT ANNOTATED\_CDS.  
 CC EMBL; U65072; AAC52831.1; ALT\_INIT.  
 CC EMBL; U65072; AAC52832.1; ALT\_INIT.  
 CC HSP; P06601; 1FJL.  
 CC TRANSFAC; T03310; -;  
 CC TRANSFAC; T04233; -;  
 CC MGD; MGI:1201673; Shox2.  
 CC InterPro; IPR003654; Homeo\_OAR.  
 CC InterPro; IPR001356; Homeobox.  
 CC InterPro; IPR000047; HTH lambdarepressor.  
 CC InterPro; IPR007104; Paired\_homeo.  
 CC Pfam; PF00046; homeobox; 1.  
 CC Pfam; PF03626; OAR; 1.  
 CC PRINTS; PR00031; HTHREPRESSOR.  
 CC ProDom; PD000010; Homeobox; 1.  
 CC SMART; SM00389; HOX; 1.  
 CC PROSITE; PS00027; HOMEBOX\_1; 1.  
 CC PROSITE; PS00071; HOMEBOX\_2; 1.  
 CC PROSITE; PS08003; OAR; 1.  
 CC Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 CC Alternative splicing.  
 CC DNA\_BIND 140 199 HOMEBOX.  
 CC DOMAIN 313 326 OAR DOMAIN.  
 CC DOMAIN 59 82 POLY-GLY.  
 CC VARSPLIC 235 246 Missing (in isoform 2).  
 CC FT /FTID=VSP 002289.  
 CC SQ SEQUENCE 331 AA; 34905 MW; D24D297E1D73A025 CRC64;  
 CC -----  
 CC Query Match 2.2%; Score 8; DB 1; Length 331;  
 CC Best Local Similarity 100.0%; Pred. No. 4;  
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC Qy 234 AGGGAGGG 241  
 CC |||||  
 CC Db 83 AGGGAGGG 90  
 CC -----  
 CC RESULT 4  
 CC GRP\_ARATH  
 CC ID GRP\_ARATH STANDARD; PRT; 349 AA.  
 CC AC P27E83; O9LSP2; -;  
 CC DT 01-AUG-1992 (Rel. 23, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Glycine-rich cell wall structural protein precursor.  
 CC GN AT3G17050 OR K14A17.12.  
 CC OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

CC NCBI\_TaxID=3702;  
 CC (1)  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=cv. Columbia;  
 CC MEDLINE=92003708; PubMed=1912511;  
 CC Quigley F., Villiot M.B., Mache R.;  
 CC "Nucleotide sequence and expression of a novel glycine-rich protein  
 CC gene from Arabidopsis thaliana.";  
 CC Plant Mol. Biol. 17:949-952(1991).  
 CC (2)  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=cv. Columbia;  
 CC MEDLINE=20277480; PubMed=10819329;  
 CC Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;  
 CC "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
 CC features of the regions of 4,504,864 bp covered by sixty P1 and TAC  
 CC clones.";  
 CC DNA Res. 7:131-135(2000).  
 CC -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).  
 CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X58338; CAA41249.1; ALT\_INIT.  
 CC EMBL; AB026636; BAA94983.1; -;  
 CC Cell wall; Structural protein; Repeat; Signal.  
 CC SIGNAL 1 23 POTENTIAL.  
 CC CHAIN 24 349 GLYCINE-RICH CELL WALL STRUCTURAL  
 CC PROTEIN.  
 CC DOMAIN 21 338 GLY-RICH.  
 CC SEQUENCE 349 AA; 25029 MW; 386ADC6817621D46 CRC64;  
 CC -----  
 CC Query Match 2.2%; Score 8; DB 1; Length 349;  
 CC Best Local Similarity 100.0%; Pred. No. 4.2;  
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC Qy 234 AGGGAGGG 241  
 CC |||||  
 CC Db 63 AGGGAGGG 70  
 CC -----  
 CC RESULT 5  
 CC GRP1\_PETHY  
 CC ID GRP1\_PETHY STANDARD; PRT; 384 AA.  
 CC AC P09789;  
 CC DT 01-MAR-1989 (Rel. 10, Created)  
 CC DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 CC DT 01-AUG-1992 (Rel. 23, Last annotation update)  
 CC DE Glycine-rich cell wall structural protein 1 precursor.  
 CC GN GRP-1.  
 CC OS Petunia hybrida (Petunia).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; lamids; Solanales; Solanaceae; Petunia.  
 CC NCBI\_TaxID=4102;  
 CC (1)  
 CC SEQUENCE FROM N.A.  
 CC Condit C.M., Meagher R.B.;  
 CC "A gene encoding a novel glycine-rich structural protein of petunia.";  
 CC Nature 323:178-181(1986).  
 CC -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).  
 CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).  
 CC -!- MISCELLANEOUS: THIS PROTEIN CONTAINS 67 GLYCINE RESIDUES.  
 CC -!- MISCELLANEOUS: 90% OF THE MATURE PROTEIN RESIDUES ARE CAPABLE OF  
 CC FORMING A BETA-PLEATED SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.  
 CC -!- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED  
 CC FAMILIES OF REPEATS, F1 AND F2, EACH REPEAT CONTAINING ABOUT 40

```

CC AA
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC DR EMBL; X04335; CAA27866.1; -
CC DR PIR; A26099;
CC DR Cell wall; Structural protein; Repeat; Signal.
CC FT SIGNAL 1 27
CC FT CHAIN 28 384 GLYCINE-RICH CELL WALL STRUCTURAL
CC FT DOMAIN 41 384 GLY-RICH.
CC FT SEQUENCE 384 AA; 28777 MW; C8541C549417D18C CRC64;
CC
CC Query Match 2.2%; Score 8; DB 1; Length 384;
CC Best Local Similarity 100.0%; Pred. No. 4.6;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 234 AGGGAGGG 241
CC DB 75 AGGGAGGG 82
CC
CC RESULT 6
CC FXD1 MOUSE STANDARD; PRT; 456 AA.
CC ID FXD1 MOUSE
CC AC Q61345;
CC DT 15-JUL-1998 (Rel. 36, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
CC DE Forkhead box protein D1 (Forkhead-related protein FXH1L8) (Forkhead-
CC related transcription factor 4) (FREAC-4).
CC GN FOXD1 OR FXH1L8 OR FREAC4 OR FHBF2.
CC OS Mus musculus (Mouse).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC OX NCBI_TaxID=10090;
CC RN [1]_TaxID=10090;
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=BALB/C; TISSUE=Embryo;
CC RX MEDLINE=95114592; PubMed=7815060;
CC RA Hatini V., Tao W., Lai E.;
CC RT "Expression of winged helix genes, BF-1 and BF-2, define adjacent
CC domains within the developing forebrain and retina.";
CC RL J. Neurobiol. 25:1293-1309(1994).
CC CC -1- FUNCTION: MAY FUNCTION IN SPECIFYING POSITIONAL IDENTITY IN THE
CC DEVELOPING RETINA AS WELL AS THE SUBDIVISION OF THE FOREBRAIN
CC NEUROEPITHELIUM.
CC CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE CNS AND
CC TEMPORAL HALF OF THE RETINA. ALSO EXPRESSED IN THE CONDENSED HEAD
CC MESENCHYME, METANEPHRIC BLASTOMA OF THE DEVELOPING KIDNEY, CORTEX
CC OF THE ADRENAL GLAND, CONDENSED MESENCHYME AT THE BASE OF THE
CC FOLLICLES OF VIBRASSAE, AND CARTILAGE PERICHONDRUM OF THE
CC DEVELOPING VERTEBRATE.
CC CC -1- DEVELOPMENTAL STAGE: AT E9.5 EMBRYOS, EXPRESSED IN A LIMITED
CC REGION OF THE NEUROEPITHELIUM AND ALSO IN THE TEMPORAL HALF OF THE
CC PRIMARY OPTIC CUP AND THE OPTIC STALK. AT E10.5, SEEN IN THE
CC HYPOTHALAMUS, TEMPORAL HALF OF THE OPTIC STALK, AND TEMPORAL
CC HEMIRETINA. AT E12.5 AND E13.5 A HIGH EXPRESSION IS SEEN IN
CC REGIONS OF CONDENSED MESENCHYME OF THE HEAD, AND AS
CC NEUROEPITHELIAL CELLS BEGIN TO DIFFERENTIATE AND MIGRATE OUTWARD
CC FROM THE VENTRICULAR ZONE, EXPRESSION DECLINES MARKEDLY. BY E16.5
CC LEVELS ARE DIMINISHED AND RESTRICTED TO UNFUSED POCKETS ALONG THE
CC EXHAUSTED VENTRICULAR ZONE.
CC CC -1- SIMILARITY: Contains 1 fork-head domain.
CC CC -1- CAUTION: WAS ORIGINALLY (REF.1) ASSIGNED TO BE BF-2 (FOXG1A).

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC DR EMBL; L38607; AAC42042.1; -
CC DR HSSP; O63245; 2HFH.
CC DR TRANSPAC; T02293; -
CC DR MGD; MGI:1347463; Foxd1.
CC DR InterPro; IPR001766; TF_Fork_head.
CC DR Pfam; PF00250; Fork_head; 1.
CC DR PRINTS; PR00053; FORKHEAD.
CC DR ProDom; PD000425; TF_Fork_head; 1.
CC DR SMART; SM00339; FH; 1
CC DR PROSITE; PS00657; FORK_HEAD_1; 1.
CC DR PROSITE; PS00658; FORK_HEAD_2; 1.
CC DR PROSITE; PS50039; FORK_HEAD_3; 1.
CC KW Transcription regulation; DNA-binding; Nuclear protein;
CC Developmental protein.
CC FT DOMAIN 29 32 POLY-GLU.
CC FT DOMAIN 33 36 POLY-ASP.
CC FT DOMAIN 38 46 POLY-GLY.
CC FT DOMAIN 55 58 POLY-ARG.
CC FT DOMAIN 73 76 POLY-ASP.
CC FT DNA_BIND 130 224 FORK-HEAD.
CC FT DOMAIN 261 267 POLY-PRO.
CC FT DOMAIN 293 302 POLY-ALA.
CC FT DOMAIN 308 319 POLY-PRO.
CC FT DOMAIN 395 400 POLY-GLY.
CC FT DOMAIN 420 425 POLY-ALA.
CC SQ SEQUENCE 456 AA; 45429 MW; 43D93F89BBDPCCC7 CRC64;
CC
CC Query Match 2.2%; Score 8; DB 1; Length 456;
CC Best Local Similarity 100.0%; Pred. No. 5.5;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 234 AGGGAGGG 241
CC DB 108 AGGGAGGG 115
CC
CC RESULT 7
CC GRP2 PHAVU STANDARD; PRT; 465 AA.
CC ID PHAVU PHAVU
CC AC P10496;
CC DT 01-JUL-1989 (Rel. 11, Created)
CC DT 01-JUL-1989 (Rel. 11, Last sequence update)
CC DT 01-OCT-1994 (Rel. 30, Last annotation update)
CC DE Glycine-rich cell wall structural protein 1.8 precursor (GRP 1.8).
CC OS Phaseolus vulgaris (Kidney bean) (French bean).
CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC OC eurosids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
CC OX NCBI_TaxID=3885;
CC RN [1]_TaxID=3885;
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=cv. Tendergreen;
CC RX MEDLINE=89091109; PubMed=3208742;
CC RA Keller B., Sauer N., Lamb C.J.;
CC RT "Glycine-rich cell wall proteins in bean: gene structure and
CC association of the protein with the vascular system.";
CC RL EMBO J. 7:3625-3633(1988).
CC CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Cell wall (Potential).
CC CC -1- DOMAIN: THE REPEATED DOMAINS OF THE PROTEIN FORM A BETA-PLATED
CC SHEET CONFIGURATION.
CC CC -1- SIMILARITY: THE N-TERMINAL SIGNAL SEQUENCE OF ABOUT 30 AA SHOWS
CC ABOUT 60% HOMOLOGY TO THAT OF THE GLYCINE-RICH CELL WALL
CC STRUCTURAL PROTEIN GRP 1.0.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X13596; CAA31932.1; --  
DR PIR; S01820;  
KW Cell wall; Structural protein; Repeat; Signal.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 465 GLYCINE-RICH CELL WALL STRUCTURAL  
FT PROTEIN 1.8.  
FT DOMAIN 33 465 GLY-RICH.  
FT DOMAIN 205 359 8 X 22 AA TANDEM REPEATS.  
FT SEQUENCE 465 AA; 36683 MW; B5C4A9B983B43607 CRC64;

Query Match 2.2%; Score 8; DB 1; Length 465;  
Best Local Similarity 100.0%; Pred. No. 5.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
| | | | |  
DB 191 AGGAGGG 198

RESULT 8  
FXD2 HUMAN STANDARD; PRT; 497 AA.  
AC O60548;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Forkhead box protein D2 (forkhead-related protein FKHL17) (Forkhead-  
DE related transcription factor 9) (FREAC-9).  
GN FOXD2 OR FKHL17 OR FREAC9.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98066765; PubMed=9403061;  
RA Ernstson S., Betz R., Lagercrantz S., Larsson C., Erickson S.,  
RA Cederberg A., Carlsson P., Enerbaeck S.;  
RT "Cloning and characterization of freac-9 (FKHL17), a novel kidney-  
RL expressed human forkhead gene that maps to chromosome 1p32-p34.";  
RL Genomics 46:78-85(1997).  
RN [2]  
RP REVISIONS  
RA Enerbaeck S.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.  
CC -1- SIMILARITY: Contains 1 fork-head domain.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF042832; AAC15421.1; --  
DR HSSP; Q63245; 2HFH.  
DR TRANSPAC; T02485;  
DR Genew; HGNC:3803; FOXD2.  
DR MM; 602211; --  
DR GO; GO:0003700; F:transcription factor activity; TAS.  
DR InterPro; IPR001766; TF\_Fork\_head.

DR Pfam; PF00250; Fork\_head; 1.  
DR PRINTS; PR00053; FORKHEAD.  
DR ProDom; PD000425; TF\_Fork\_head; 1.  
DR SMART; SM00339; FH; 1.  
DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE; PS00339; FORK\_HEAD\_3; 1.  
KW DNA-binding; Nuclear protein; Transcription regulation.  
FT DOMAIN 90 94 POLY-ALA.  
FT DOMAIN 101 104 POLY-ALA.  
FT DNA\_BIND 126 217 FORK-HEAD.  
FT DOMAIN 247 250 POLY-ALA.  
FT DOMAIN 296 306 POLY-ALA.  
FT DOMAIN 398 409 POLY-GLY.  
FT DOMAIN 421 426 POLY-GLY.  
FT DOMAIN 442 445 POLY-ALA.  
SQ SEQUENCE 497 AA; 49007 MW; EAAF498D216BE019 CRC64;

Query Match 2.2%; Score 8; DB 1; Length 497;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
| | | | |  
DB 397 AGGAGGG 404

RESULT 9  
IRBP MOUSE STANDARD; PRT; 610 AA.  
AC P49194; Q9ROH8;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Interphotoreceptor retinoid-binding protein precursor (IRBP)  
DE (interstitial retinol-binding protein) (Fragment).  
GN RBP3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94115692; PubMed=1342928;  
RA Stanhope M.J., Czelusniak J., Si J.-S., Nickerson J., Goodman M.;  
RT "A molecular perspective on mammalian evolution from the gene encoding  
RT interphotoreceptor retinoid binding protein, with convincing evidence  
RL for bat monophyly.";  
RL Mol. Phylogenet. Evol. 1:148-160(1992).  
RN [3]  
RP SEQUENCE OF 71-454 FROM N.A.  
RC STRAIN=MOA;  
RA Suzuki H., Serizawa K.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: IRBP SHUTTLES 11-CIS AND ALL TRANS RETINIDS BETWEEN  
CC THE RETINOL ISOMERASE IN THE PIGMENT EPITHELIUM AND THE VISUAL  
CC PIGMENTS IN THE PHOTORECEPTOR CELLS OF THE RETINA.  
CC -1- SUBCELLULAR LOCATION: INTERPHOTORECEPTOR MATRIX THAT PERMEATES  
CC THE SPACE BETWEEN THE RETINA AND THE CONTIGUOUS LAYER OF PIGMENT  
CC EPITHELIUM CELLS.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PHOTORENSITIVE TISSUES;  
CC RETINA AND PINEAL GLAND.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF126968; AAA39331.2; --  
 DR EMBL; AB033711; BAA85872.1; --  
 DR MGD; MGI:97878; Bdp3.  
 DR InterPro; IPR003581; TSPC.  
 DR Pfam; PF02692; IRBP; 2.  
 DR SMART; SM00245; TSPC; 2.  
 KW Vitamin A; Transport; Repeat; Signal; Glycoprotein.  
 FT SIGNAL 1 17  
 FT CHAIN 18 >610  
 FT DOMAIN 18 >610  
 FT REPEAT 18 320  
 FT REPEAT 321 >610  
 FT CARBOHYD 107 >610  
 FT CARBOHYD 205 205  
 FT CARBOHYD 513 513  
 FT CONFLICT 161 161  
 FT NON\_TER 610 610  
 SQ SEQUENCE 610 AA; 66679 MW; F392FE2D187C6416 CRC64;

Query Match 2.2%; Score 8; DB 1; Length 610;  
 Best Local Similarity 100.0%; Pred. No. 7.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 102 LEAPQAP 109  
 DB 96 LEAPQAP 103

## RESULT 10

YC1Q\_ECOLI  
 ID YC1Q\_ECOLI STANDARD; PRT; 631 AA.  
 AC P45848; P45849; P76031;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yciQ.  
 GN YC1Q OR B1268.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RA Milkman R., McKane M.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MGI655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Alba H., Baba T., Fujita K., Hayaashi K., Inada T., Isono K., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshina T., Saito N.,  
 RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
 RA Takenoko K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
 RT "A 570-Kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map.";

RL DNA Res. 3:363-377(1996).  
 CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 CC FRAMESHIFT IN POSITION 198.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U18111; AAB59988.1; ALT\_FRAME.  
 DR EMBL; U18111; AAB59989.1; ALT\_FRAME.  
 DR EMBL; AB000224; AAC74350.1; --  
 DR EMBL; D90764; BAA14803.1; ALT\_INIT.  
 DR EMBL; D90765; BAA14818.1; ALT\_INIT.  
 DR PIR; G64874; G64874.  
 DR EcGene; EG12969; yciQ.  
 KW Hypothetical protein; Complete proteome.  
 FT DOMAIN 616 630 POLY-GLY.  
 FT CONFLICT 84 84 A -> G (IN REF. 1).  
 FT CONFLICT 88 88 N -> D (IN REF. 1).  
 SQ SEQUENCE 631 AA; 71289 MW; 1154FF72FDC6C16E CRC64;

Query Match 2.2%; Score 8; DB 1; Length 631;  
 Best Local Similarity 100.0%; Pred. No. 7.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 234 AGGGAGGG 241  
 DB 619 AGGGAGGG 626

## RESULT 11

CADM\_HUMAN  
 ID CADM\_HUMAN STANDARD; PRT; 828 AA.  
 AC Q9UJ99; O43205;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cadherin-22 precursor.  
 GN CDH22 OR C20ORF25.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,  
 RA Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M., B.J.,  
 RA Beasley O.P., Bird C.P., Blake S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lehaesalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A.C., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,

RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871 (2001).  
 RN [2]  
 RP SEQUENCE OF 449-828 FROM N.A.  
 RC TISSUE=Brain;  
 RA Yu W., Sarginson J., Gibbs R.A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.  
 CC They preferentially interact with themselves in a homophilic  
 CC manner in connecting cells; cadherins may thus contribute to the  
 CC sorting of heterogeneous cell types. PB-cadherins may have a role  
 CC in the morphological organization of pituitary gland and brain  
 CC tissues (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: Contains 5 cadherin domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC ENBL; AL031687; CAB51587.2; -  
 CC EMBL; AF035300; AAB88183.1; -  
 CC Genew; HGNC:13251; CDH22.  
 CC HSSP; P15116; INCUJ.  
 CC InterPro; IPR002126; Cadherin.  
 CC InterPro; IPR002233; Cadherin\_C\_term.  
 CC Pfam; PF00028; Cadherin; 5.  
 CC Pfam; PF01049; Cadherin\_C\_term; 1.  
 CC PRINTS; P00205; CADHERIN.  
 CC SMART; SM00112; CA; 5.  
 CC PROSITE; PS00212; CADHERIN 1; 2.  
 CC PROSITE; PS0268; CADHERIN 2; 5.  
 CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Signal.  
 KW SIGNAL 1 34 POTENTIAL.  
 FT CHAIN 35 828 CADHERIN-22.  
 FT DOMAIN 36 624 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 625 645 POTENTIAL.  
 FT DOMAIN 646 828 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1 64 168 CADHERIN 1.  
 FT DOMAIN 2 169 277 CADHERIN 2.  
 FT DOMAIN 3 278 394 CADHERIN 3.  
 FT DOMAIN 4 395 498 CADHERIN 4.  
 FT DOMAIN 5 499 616 CADHERIN 5.  
 FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 828 AA; 89091 MW; 520F7B1DF624DCA CRC64;  
 Query Match 2.2%; Score 8; DB 1; Length 828;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 234 AGGGAGGG 241  
 |||||  
 Db 706 AGGGAGGG 713  
 RESULT 12  
 KDGE DROME  
 ID\_KDGE DROME STANDARD; PRT; 1454 AA.  
 AC Q09103;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Eye-specific diacylglycerol kinase (EC 2.7.1.107) (Retinal  
 DE degeneration A protein) (Diglyceride kinase 2) (DGK 2) (DAG kinase 2).

GN RDGA OR DGK2.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Canon-S; TISSUE=Head;  
 RX MEDLINE=9468563; PubMed=8248222;  
 RA Masai I., Okazaki A., Hosoya T., Hotta Y.;  
 RT "Drosophila retinal degeneration A gene encodes an eye-specific  
 RT diacylglycerol kinase with cysteine-rich zinc-finger motifs and  
 RT ankyrin repeats.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:11157-11161 (1993).  
 CC -1- FUNCTION: REQUIRED FOR THE MAINTENANCE OF THE PHOTORECEPTOR. ITS  
 CC ABSENCE LEADS TO RHADOMERE DEGENERATION DUE TO DEFECTIVE  
 CC PHOSPHOLIPID TURNOVER.  
 CC -1- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-  
 CC diacylglycerol 3-phosphate.  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT EYE.  
 CC -1- DISEASE: MUTANTS OF THIS GENE ARE CHARACTERIZED BY HAVING  
 CC PHOTORECEPTOR CELLS THAT DEVELOP NORMALLY BUT DEGENERATE RAPIDLY  
 CC AFTER ECLOSION.  
 CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE  
 CC FAMILY.  
 CC -1- SIMILARITY: Contains 2 zinc-dependent phorbol-ester and DAG  
 CC binding domains.  
 CC -1- SIMILARITY: Contains 4 ANK repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC ENBL; D17315; BAA04135.1; -  
 CC PIR; T13709; T13709.  
 CC FlyBase; FBgn0003217; rdga.  
 CC GO; GO:0016059; P:deactivation of rhodopsin mediated signaling; IGI.  
 CC GO; GO:0016056; P:rhodopsin mediated signaling; IMP.  
 CC InterPro; IPR002110; ANK.  
 CC InterPro; IPR002219; DAG\_PE-bind.  
 CC InterPro; IPR00756; DAGKa.  
 CC InterPro; IPR001206; DAGKc.  
 CC Pfam; PF00023; ank; 4.  
 CC Pfam; PF00130; DAG\_PE-bind; 1.  
 CC Pfam; PF00609; DAGKa; 1.  
 CC Pfam; PF00781; DAGKc; 1.  
 CC ProDom; PD002939; DAGKa; 1.  
 CC ProDom; PD005043; DAGKc; 1.  
 CC SMART; SM00248; ANK; 2.  
 CC SMART; SM00109; C1; 2.  
 CC SMART; SM00045; DAGKa; 1.  
 CC SMART; SM00046; DAGKc; 1.  
 CC PROSITE; PS50088; ANK REPEAT; 2.  
 CC PROSITE; PS00297; ANK REP REGION; 1.  
 CC PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; FALSE NEG.  
 CC PROSITE; PS50081; DAG\_PE\_BIND\_DOM\_2; FALSE NEG.  
 KW transferase; Kinase; ANK repeat; Repeat; Vision;  
 KW Phorbol-ester binding.  
 FT DOMAIN 592 642 PHORBOL-ESTER AND DAG BINDING 1.  
 FT DOMAIN 662 719 PHORBOL-ESTER AND DAG BINDING 2.  
 FT DOMAIN 807 935 CATALYTIC-A (POTENTIAL).  
 FT DOMAIN 961 1115 CATALYTIC-B (POTENTIAL).  
 FT REPEAT 1317 1346 ANK 1.  
 FT REPEAT 1350 1379 ANK 2.  
 FT REPEAT 1386 1415 ANK 3.  
 FT REPEAT 1419 1448 ANK 4.  
 FT DOMAIN 2 6 POLY-GLN.

```

FT DOMAIN 24 39 THR-RICH.
FT DOMAIN 110 115 POLY-SER.
FT DOMAIN 227 231 POLY-GLU.
FT DOMAIN 758 775 GLY-RICH.
FT MUTAGEN 869 869 G->D: IN MUTANT RDGA2.
SQ SEQUENCE 1454 AA; 159675 MW; C84C81095PEA16AA CRC64;

Query Match 2.2%; Score 8; DB 1; Length 1454;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241
Db 764 AGGAGGG 771

RESULT 13
NEUG_BOVIN
ID NEUG_BOVIN STANDARD; PRT; 76 AA.
AC F35722;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogranin (NG) (P17) (B-50 immunoreactive C-kinase substrate)
DE (BICKS) (Fragment).
GN NRGN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP MEDLINE=91093129; PubMed=1824695;
RA Baudier J., Deloume J.C., van Dorsselaer A., Black D.,
RT "Purification and characterization of a brain-specific protein kinase
RT C substrate, neurogranin (p17). Identification of a consensus amino
RT acid sequence between neurogranin and neuromodulin (GAP43) that
RT corresponds to the protein kinase C phosphorylation site and the
RT calmodulin-binding domain."
RL J. Biol. Chem. 266:229-237(1991).
CC -!- FUNCTION: ACTS AS A "THIRD MESSENGER" SUBSTRATE OF PROTEIN KINASE
CC C-MEDIATED MOLECULAR CASCADES DURING SYNAPTIC DEVELOPMENT AND
CC REMODELING. BINDS TO CALMODULIN IN THE ABSENCE OF CALCIUM.
CC -!- TISSUE SPECIFICITY: IS HIGHLY ENRICHED IN BRAIN ACCUMULATES
CC POSTSYNAPTICALLY IN DENDRITIC SPINES OF NEOSTRIATAL NEURONS.
CC -!- MASS SPECTROMETRY: MW=7837.1; MW ERR=0.5; METHOD=Electrospray.
CC -!- SIMILARITY: BELONGS TO THE NEUROGRANIN FAMILY.
CC PIR; A39034; A39034.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00612; IQ; 1.
DR SMART; SM00015; IQ; 1.
DR PROSITE; PS50096; IQ; 1.
KW Calmodulin-binding; Phosphorylation; Neurone.
FT DOMAIN 24 47 IQ.
FT MOD_RES 48 76 COLLAGEN-LIKE.
FT MOD_RES 34 34 PHOSPHORYLATION (BY PKC).
FT NON_TER 76 76
SQ SEQUENCE 76 AA; 7248 MW; DA898C93FD6165CF CRC64;

Query Match 1.9%; Score 7; DB 1; Length 76;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGP 242
Db 67 GGAGGGP 73

RESULT 14
NEUG_BOVIN
ID NEUG_BOVIN STANDARD; PRT; 76 AA.
AC F35722;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogranin (NG) (P17) (B-50 immunoreactive C-kinase substrate)
DE (BICKS) (Fragment).
GN NRGN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP MEDLINE=91093129; PubMed=1824695;
RA Baudier J., Deloume J.C., van Dorsselaer A., Black D.,
RT "Purification and characterization of a brain-specific protein kinase
RT C substrate, neurogranin (p17). Identification of a consensus amino
RT acid sequence between neurogranin and neuromodulin (GAP43) that
RT corresponds to the protein kinase C phosphorylation site and the
RT calmodulin-binding domain."
RL J. Biol. Chem. 266:229-237(1991).
CC -!- FUNCTION: ACTS AS A "THIRD MESSENGER" SUBSTRATE OF PROTEIN KINASE
CC C-MEDIATED MOLECULAR CASCADES DURING SYNAPTIC DEVELOPMENT AND
CC REMODELING. BINDS TO CALMODULIN IN THE ABSENCE OF CALCIUM.
CC -!- TISSUE SPECIFICITY: IS HIGHLY ENRICHED IN BRAIN ACCUMULATES
CC POSTSYNAPTICALLY IN DENDRITIC SPINES OF NEOSTRIATAL NEURONS.
CC -!- MASS SPECTROMETRY: MW=7837.1; MW ERR=0.5; METHOD=Electrospray.
CC -!- SIMILARITY: BELONGS TO THE NEUROGRANIN FAMILY.
CC PIR; A39034; A39034.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00612; IQ; 1.
DR SMART; SM00015; IQ; 1.
DR PROSITE; PS50096; IQ; 1.
KW Calmodulin-binding; Phosphorylation; Neurone.
FT DOMAIN 24 47 IQ.
FT MOD_RES 48 76 COLLAGEN-LIKE.
FT MOD_RES 34 34 PHOSPHORYLATION (BY PKC).
FT NON_TER 76 76
SQ SEQUENCE 76 AA; 7248 MW; DA898C93FD6165CF CRC64;

Query Match 1.9%; Score 7; DB 1; Length 76;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGP 242
Db 67 GGAGGGP 73

RESULT 15
NEUG_HUMAN
ID NEUG_HUMAN STANDARD; PRT; 78 AA.
AC Q92C66;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Neurogranin (NG) (RC3).
GN NRGN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=97288522; PubMed=9143500;
RA Martinez de Arrieta C., Perez Jurado L., Bernal J., Coloma A.;

```

```

NEUG_CAPHI
ID NEUG_CAPHI STANDARD; PRT; 78 AA.
AC P54877;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogranin (NG) (Protein kinase C substrate 7.5 kDa protein) (RC3).
GN NRGN.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95287801; PubMed=7539519;
RA Piosik P.A., van Groenigen M., Ponne N.J., Bolhuis P.A., Baas F.;
RT "RC3/neurogranin structure and expression in the caprine brain in
RT relation to congenital hypothyroidism."
RL Brain Res. Mol. Brain Res. 29:119-130(1995).
CC -!- FUNCTION: ACTS AS A "THIRD MESSENGER" SUBSTRATE OF PROTEIN KINASE
CC C-MEDIATED MOLECULAR CASCADES DURING SYNAPTIC DEVELOPMENT AND
CC REMODELING. BINDS TO CALMODULIN IN THE ABSENCE OF CALCIUM (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NEUROGRANIN FAMILY.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; S78295; AAB34481.1; -.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00612; IQ; 1.
DR SMART; SM00015; IQ; 1.
DR PROSITE; PS50096; IQ; 1.
KW Calmodulin-binding; Phosphorylation; Neurone.
FT DOMAIN 26 47 IQ.
FT MOD_RES 48 78 COLLAGEN-LIKE.
FT MOD_RES 36 36 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
SQ SEQUENCE 78 AA; 7549 MW; 8E47CDA39F085794 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGP 242
Db 69 GGAGGGP 75

RESULT 15
NEUG_HUMAN
ID NEUG_HUMAN STANDARD; PRT; 78 AA.
AC Q92C66;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Neurogranin (NG) (RC3).
GN NRGN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=97288522; PubMed=9143500;
RA Martinez de Arrieta C., Perez Jurado L., Bernal J., Coloma A.;

```





CC POSTSYNAPTICALLY IN DENDRITIC SPINES OF NEOSTRIATAL NEURONS.  
CC -1- PTM: PHOSPHORYLATION IS ACTIVATED BY CALCIUM, PHOSPHOLIPID, AND  
CC DIACYLGLYCEROL. PHOSPHORYLATION INHIBITS BINDING TO CALMODULIN  
CC BOTH IN THE PRESENCE AND ABSENCE OF CALCIUM.  
CC -1- SIMILARITY: BELONGS TO THE NEUROGRANIN FAMILY.  
CC -1- SIMILARITY: Contains 1 IQ domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L09119; AAA42023.1; -;  
CC EMBL; U22062; AAA80223.1; -;  
CC PIR; A57288; A57288.  
CC InterPro; IPR000048; IQ\_region.  
CC Pfam; PF00612; IQ; 1.  
CC SMART; SM00015; IQ; 1.  
CC PROSITE; P850096; IQ; 1.  
CC Calmodulin-binding; Phosphorylation; Neurone.  
FT MOD\_RES 1 1 BLOCKED.  
FT DOMAIN 26 47 IQ.  
FT DOMAIN 48 78 COLLAGEN-LIKE.  
FT MOD\_RES 36 36 PHOSPHORYLATION (BY PKC).  
SQ SEQUENCE 78 AA; 7496 MW; 8E47CDB38E095794 CRC64;  
  
Query Match 1.9%; Score 7; DB 1; Length 78;  
Best Local Similarity 100.0%; Pred. No. 9.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 236 GGAGGGP 242  
DB 69 GGAGGGP 75  
  
RESULT 17  
Y1B3\_CLOAB  
ID Y1B3\_CLOAB STANDARD; PRT; 96 AA.  
AC P34159;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein CAC3713.  
GN CAC3713.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]\_TaxID=1488;  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=93273706; PubMed=8501044;  
RA Sauer U., Duerre P.;  
RT "Sequence and molecular characterization of a DNA region encoding a  
RT small heat shock protein of Clostridium acetobutylicum.";  
RL J. Bacteriol. 175:3394-3400(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R.L., Lee H.W., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
RT bacterium Clostridium acetobutylicum.";  
RL J. Bacteriol. 183:4823-4838(2001).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X65276; CAA46375.1; -;  
CC EMBL; AE007866; AAK81633.1; -;  
CC PIR; F97355; F97355.  
CC KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 96 AA; 11003 MW; 0E427F533D1D79DF CRC64;  
  
Query Match 1.9%; Score 7; DB 1; Length 96;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 319 ISVTPEE 325  
DB 4 ISVTPEE 10  
  
RESULT 18  
NLPD\_YEREN  
ID NLPD\_YEREN STANDARD; PRT; 97 AA.  
AC P47764;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Lipoprotein nlpD (Fragment).  
GN NLPD.  
OS Versinia enterocolitica.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Versinia.  
OX NCBI\_TaxID=630;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=W1024 / Serotype O:9;  
RX MEDLINE=95247270; PubMed=7729893;  
RA Iriarte M., Stainier I., Cornelis G.R.;  
RT "The rps gene from Versinia enterocolitica and its influence on  
RT expression of virulence factors.";  
RL Infect. Immun. 63:1840-1847(1995).  
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid  
CC anchor (Potential).  
CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U16152; AAC43390.1; -;  
CC MEROPS; M37.UPW; -;  
DR InterPro; IPR002886; Peptidase M37.  
DR InterPro; IPR000437; Prok\_lipoProt.  
DR Pfam; PF01551; Peptidase\_M37; 1.  
DR PROSITE; PS00013; PROKAR LIPOPROTEIN; PARTIAL.  
KW Inner membrane; Lipoprotein.  
FT NON\_TER 1  
RP SEQUENCE 97 AA; 10667 MW; 9E194F9A13D0838E CRC64;  
  
Query Match 1.9%; Score 7; DB 1; Length 97;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 81 GSTGTSS 87  
DB 68 GSTGTSS 74



RESULT 19  
PER\_DROER  
ID PER\_DROER STANDARD; PRT; 110 AA.  
AC Q26288;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Period circadian protein (Fragment).  
GN PER.  
OS Drosophila erecta (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7220;  
RN [1]  
RP SEQUENCE FROM N.A.; PubMed=1487825;  
RX MEDLINE=93140158; PubMed=1487825;  
RA Peixoto A.A., Costa R., Wheeler D.A., Hall J.C., Kyriacou C.P.;  
RT "Evolution of the threonine-glycine repeat region of the period gene  
in the melanogaster species subgroup of Drosophila.";  
RL J. Mol. Evol. 35:411-419(1992).  
CC -!- FUNCTION: ESSENTIAL FOR BIOLOGICAL CLOCK FUNCTIONS. DETERMINES THE  
PERIOD LENGTH OF CIRCADIAN AND ULTRADIAN RHYTHMS; AN INCREASE IN  
PER DOSAGE LEADS TO SHORTENED CIRCADIAN RHYTHMS AND A DECREASE  
LEADS TO LENGTHENED CIRCADIAN RHYTHMS. ESSENTIAL FOR THE CIRCADIAN  
RHYTHMICITY OF LOCOMOTOR ACTIVITY, ECLOSION BEHAVIOR, AND FOR THE  
RHYTHMIC COMPONENT OF THE MALE COURTSHIP SONG THAT ORIGINATES IN  
THE THORACIC NERVOUS SYSTEM. THE BIOLOGICAL CYCLE DEPENDS ON THE  
RHYTHMIC FORMATION AND NUCLEAR LOCALIZATION OF THE TIM-PER  
COMPLEX. LIGHT INDUCES THE DEGRADATION OF TIM, WHICH PROMOTES  
ELIMINATION OF PER. NUCLEAR ACTIVITY OF THE HETERODIMER  
COORDINATIVELY REGULATES PER AND TIM TRANSCRIPTION THROUGH A  
NEGATIVE FEEDBACK LOOP. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN  
TRANSCRIPTIONAL LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING  
INDIRECT TRANSCRIPTIONAL INHIBITION (BY SIMILARITY).  
CC -!- SUBUNIT: FORMS HETERODIMER WITH TIMELESS (TIM); THE COMPLEX THEN  
TRANSLLOCATES INTO THE NUCLEUS (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR AT SPECIFIC PERIODS OF THE DAY.  
CC -!- FIRST ACCUMULATES IN THE PERINUCLEAR REGION ABOUT ONE HOUR BEFORE  
TRANSLLOCATION INTO THE NUCLEUS. INTERACTION WITH TIM IS REQUIRED  
FOR NUCLEAR LOCALIZATION (BY SIMILARITY).  
CC -!- DOMAIN: THE RUN OF GLY-THR IS IMPLICATED IN THE MAINTENANCE OF  
CIRCADIAN PERIOD AT DIFFERENT TEMPERATURES. DELETION OF THE REPEAT  
LEADS TO A SHORTENING OF THE COURTSHIP SONG CYCLE PERIOD, AND THUS  
COULD BE IMPORTANT FOR DETERMINING FEATURES OF SPECIES-SPECIFIC  
MATING BEHAVIOR (BY SIMILARITY).  
CC -!- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE  
DOUBLE-TIME PROTEIN (DET). PHOSPHORYLATION COULD BE IMPLICATED IN  
THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER  
PER-TIM (BY SIMILARITY).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; S53301; AAB25031.2; .  
DR FlyBase; FBgn0012277; Dere\per.  
KW Biological rhythms; Repeat; Nuclear protein; Phosphorylation.  
FT NON\_TER 1  
FT DOMAIN 33 64 G-T REPEATS.  
FT DOMAIN 85 95 POLY-GLY.  
FT NON\_TER 110 110  
SQ SEQUENCE 110 AA; 9850 MW; 2BAA187EF55CDA95 CRC64;  
Query Match 1.9%; Score 7; DB 1; Length 110;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 235 GGGAGGG 241

Db 87 GGGAGGG 93  
|||||  
RESULT 20  
HBB\_DASAK  
ID HBB\_DASAK STANDARD; PRT; 141 AA.  
AC P56692;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hemoglobin beta chain.  
GN HBB.  
OS Dasyatis akajei (Red stingray) (Akaii).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Squalia; Hynosaquala; Pristionotidae; Batoidea;  
OC Myliobatiformes; Myliobatoidei; Dasyatidae; Dasyatis.  
OX NCBI\_TaxID=31902;  
RN [1]  
RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
RX TISSUE=Blood;  
RX MEDLINE=99322407; PubMed=10393295;  
RA Chong K.T., Miyazaki G., Morimoto H., Oda Y., Park S.-Y.;  
RT "Structures of the deoxy and CO forms of haemoglobin from Dasyatis  
akajei, a cartilaginous fish.";  
RL Acta Crystallogr. D 55:1291-1300(1999).  
CC -!- FUNCTION: INVOLVED IN OXYGEN TRANSPORT FROM THE LUNG TO THE  
VARIOUS PERIPHERAL TISSUES.  
CC -!- SUBUNIT: Heterotetramer of two alpha chains and two beta chains.  
CC -!- TISSUE SPECIFICITY: Red blood cells.  
CC -!- SIMILARITY: Belongs to the globin family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB023723; BAA75250.1; .  
DR PDB; 1CG5; 01-APR-99.  
DR PDB; 1CG8; 01-APR-99.  
DR InterPro; IPR000971; Globin.  
DR Pfam; PF00042; globin; 1.  
DR PROSITE; PS01033; GLOBIN; 1.  
KW Heme; Oxygen transport; Transport; Erythrocyte;  
KW 3D-structure.  
FT INIT\_MET 0 0  
FT METAL 59 59 IRON (HEME DISTAL LIGAND).  
FT METAL 88 88 IRON (HEME PROXIMAL LIGAND).  
FT HELIX 5 17  
FT HELIX 20 34  
FT HELIX 36 39  
FT TURN 40 41  
FT TURN 43 45  
FT TURN 46 47  
FT TURN 51 52  
FT HELIX 54 72  
FT TURN 73 75  
FT HELIX 77 80  
FT TURN 81 81  
FT HELIX 82 91  
FT TURN 92 92  
FT TURN 95 95  
FT HELIX 96 113  
FT HELIX 114 116  
FT HELIX 119 137  
FT TURN 138 139  
SQ SEQUENCE 141 AA; 16289 MW; DAED4F578904D27B CRC64;  
Query Match 1.9%; Score 7; DB 1; Length 141;  
Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 DKVQAL 169  
 Db 61 DKVQAL 67

RESULT 21  
 TWIST\_HUMAN STANDARD; PRT; 202 AA.  
 AC Q15672; Q92487; Q99804;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Twist related protein 1 (H-twist).  
 GN TWIST1 OR TWIST.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC TISSUE=Lung;  
 RA Wang S.M., Pignolo R.J., Rotenberg M.O., Coljee V.W.,  
 RA Cristofalo V.J., Sierra F.;  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=97148940; PubMed=8995765;  
 RA Bourgeois P., Stoetzel C., Bolcato-Bellemin A.-L., Mattei M.-G.,  
 RA Perrin-Schmitt F.;  
 RT "The human H-twist gene is located at 7p21 and encodes a B-HLH  
 RT protein that is 96% similar to its murine M-twist counterpart.";  
 RL Mamm. Genome 7:915-917(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND VARIANTS SCS PRO-119, ILE-135 INS AND PRO-139  
 RP INS.  
 RX MEDLINE=97141916; PubMed=8988166;  
 RA Howard T.D., Paznekas W.A., Green E.D., Chiang L.C., Ma N.,  
 RA Ortiz de Luna R.I., Delgado C.G., Gonzalez-Ramos M., Kline A.D.,  
 RA Jabs E.W.;  
 RT "Mutations in TWIST, a basic helix-loop-helix transcription factor,  
 RT in Saethre-Chotzen syndrome.";  
 RL Nat. Genet. 15:36-41(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97358582; PubMed=9215678;  
 RA Krebs I., Weis I., Hudler M., Rommens J.M., Roth H., Scherer S.W.,  
 RA Tsui L.-C., Fuchtbauer E.-M., Grzeschik K.-H., Tsuji K., Kunz J.;  
 RT "Translocation breakpoint maps 5 kb 3-prime from TWIST in a patient  
 RT affected with Saethre-Chotzen syndrome.";  
 RL Hum. Mol. Genet. 6:1079-1086(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Strong C., Graves T., Duckels G., Ozersky P.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klautner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schafer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7]  
 RN VARIANTS SCS PRO-131 AND PRO-139 INS.  
 RX MEDLINE=97141917; PubMed=8988167;  
 RA el Ghourzi V., le Merrer M., Perrin-Schmitt F., Lajeunie E.,  
 RA Benit P., Renier D., Bourgeois P., Bolcato-Bellemin A.-L.,  
 RA Munnich A., Bonaventure J.;  
 RT "Mutations of the TWIST gene in the Saethre-Chotzen syndrome.";  
 RL Nat. Genet. 15:42-46(1997).  
 RN [8]  
 RP VARIANT BGS VAL-156.  
 RX MEDLINE=21621049; PubMed=11754069;  
 RA Seto M.L., Lee S.J., Sze R.W., Cunningham M.L.;  
 RT "Another TWIST on Baller-Gerold syndrome.";  
 RL Am. J. Med. Genet. 104:323-330(2001).  
 CC [1-] FUNCTION: PROBABLE TRANSCRIPTION FACTOR, WHICH SEEMS TO BE  
 CC INVOLVED IN THE NEGATIVE REGULATION OF CELLULAR DETERMINATION AND  
 CC IN THE DIFFERENTIATION OF SEVERAL LINEAGES INCLUDING MYOGENESIS,  
 CC OSTEOGENESIS, AND NEUROGENESIS. INHIBITS MYOGENESIS BY  
 CC SEQUESTERING E PROTEINS, INHIBITING TRANS-ACTIVATION BY MEF2, AND  
 CC INHIBITING DNA-BINDING BY MYOD THROUGH PHYSICAL INTERACTION. THIS  
 CC INTERACTION PROBABLY INVOLVES THE BASIC DOMAINS OF BOTH PROTEINS  
 CC (BY SIMILARITY).  
 CC [2-] SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HOMODIMER (BY SIMILARITY).  
 CC [3-] SUBCELLULAR LOCATION: Nuclear.  
 CC [4-] TISSUE SPECIFICITY: SUBSET OF MESODERMAL CELLS.  
 CC [5-] DISEASE: Defects in TWIST1 are one of the causes of Saethre-  
 CC Chotzen syndrome (SCS) [MIM:101400]; also known as  
 CC acrocephalosyndactyly type III (ACS3). SCS is an autosomal  
 CC dominant defect characterized by minor skull and limb anomalies.  
 CC [6-] DISEASE: Defects in TWIST1 are the cause of Robinow-Sorauf  
 CC syndrome (RSS) [MIM:180750]; also known as craniosynostosis-bifid  
 CC hallux syndrome. RSS is an autosomal dominant defect characterized  
 CC by minor skull and limb anomalies which is very similar to  
 CC Saethre-Chotzen syndrome.  
 CC [7-] DISEASE: Defects in TWIST1 are a cause of Baller-Gerold syndrome  
 CC (BGS) [MIM:218600]. BGS is a craniosynostosis with radial defects  
 CC along with other features, including small, round ears with  
 CC prominent crus helices and cervical anomalies. It seems an  
 CC heterogeneous form of Saethre-Chotzen syndrome.  
 CC [8-] SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X91662; CAA62850.1; -;  
 CC EMBL; X99268; CAA67664.1; -;  
 CC EMBL; U80998; AAC50930.1; -;  
 CC EMBL; Y10871; CAA71821.1; -;  
 CC EMBL; AC003986; AAC60381.1; -;  
 CC EMBL; BC036704; AAH36704.1; -;  
 CC FIR; G01204; G01204.  
 CC TRANSFAC; T04913; -;  
 CC Genew; HGNC:12428; TWIST1.  
 CC MIM; 601622; -;  
 CC MIM; 101400; -;  
 CC MIM; 180750; -;  
 CC MIM; 218600; -;  
 CC GO; GO:0004857; F:enzyme inhibitor activity; TAS.

DR GO; GO:0005515; P:protein binding activity; TAS.  
 DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. . . ; TAS.  
 DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.  
 DR GO; GO:0000142; P:negative regulation of transcription from P. . . ; TAS.  
 DR GO; GO:0001501; P:skeletal development; TAS.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR Pfam; PF00010; HLH; 1.  
 DR SMART; SM00353; HLH; 1.  
 DR PROSITE; PS00038; HLH\_1; 1.  
 DR PROSITE; PS00888; HLH\_2; 1.  
 KW Differentiation; Developmental protein; Nuclear protein; DNA-binding;  
 KW Transcription regulation; Disease mutation; Polymorphism.  
 FT DOMAIN 80 98 GLY-RICH.  
 FT DNA BIND 108 120 BASIC DOMAIN.  
 FT DOMAIN 121 160 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT VARIANT 31 31 S -> G (in dbSNP:1800126).  
 FT VARIANT 84 84 G -> S (in dbSNP:2234705).  
 FT VARIANT 119 119 Q -> P (in SCS).  
 FT VARIANT 131 131 L -> P (in SCS).  
 FT VARIANT 135 135 I -> IAAERKII (in SCS).  
 FT VARIANT 139 139 P -> PKIITLP (in SCS).  
 FT VARIANT 156 156 I -> V (in BGS).  
 FT CONFLICT 32 32 G -> A (IN REF. 2).  
 FT CONFLICT 36 36 G -> A (IN REF. 2).  
 FT CONFLICT 41 41 S -> T (IN REF. 1 AND 4).  
 FT CONFLICT 45 45 S -> T (IN REF. 1 AND 4).  
 FT CONFLICT 56 56 MISSING (IN REF. 1 AND 4).  
 FT CONFLICT 59 59 G -> A (IN REF. 2).  
 FT CONFLICT 92 92 G -> GGGGG (IN REF. 2).  
 SQ SEQUENCE 202 AA; 20954 MW; 9394E4351BA1D081 CRC64;  
 Query Match 1.9%; Score 7; DB 1; Length 202;  
 Best Local Similarity 100.08; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 235 GGGAGGG 241  
 Db 84 GGGAGGG 90  
 RESULT 22  
 TWIS1 MOUSE STANDARD; PRT; 206 AA.  
 AC P26687;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Twist related protein 1 (M-twist).  
 GN TWIS1 OR TWIST.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91122450; PubMed=1840517;  
 RA Wolf C., Thiese C., Stoetzel C., Thiese B., Gerlinger P.,  
 RA Perrin-Schmitt F.;  
 RT "The M-twist gene of Mus is expressed in subsets of mesodermal cells  
 RT and is closely related to the Xenopus X-twi and the Drosophila twist  
 RT genes.";  
 RL Dev. Biol. 143:363-373 (1991).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=FVB/N;  
 RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Baha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 [3]  
 FUNCTION:  
 MEDLINE=98001585; PubMed=9343420;  
 RA Hanamori Y., Wu H.Y., Sartorelli V., Kedes L.;  
 RT "The basic domain of myogenic basic helix-loop-helix (bHLH) proteins  
 RT is the novel target for direct inhibition by another bHLH protein,  
 RT Twist.";  
 RL Mol. Cell. Biol. 17:6563-6573 (1997).  
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR, WHICH SEEMS TO BE  
 CC INVOLVED IN THE NEGATIVE REGULATION OF CELLULAR DETERMINATION AND  
 CC IN THE DIFFERENTIATION OF SEVERAL LINEAGES INCLUDING MYOGENESIS,  
 CC OSTEOGENESIS, AND NEUROGENESIS. INHIBITS MYOGENESIS BY  
 CC SEQUESTERING E PROTEINS, INHIBITING TRANS-ACTIVATION BY MEF2, AND  
 CC INHIBITING DNA-BINDING BY MYOD THROUGH PHYSICAL INTERACTION. THIS  
 CC INTERACTION PROBABLY INVOLVES THE BASIC DOMAINS OF BOTH PROTEINS.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC bHLH PROTEIN. HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: SUBSET OF MESODERMAL CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; M63649; AAA40514.1; -;  
 CC EMBL; M63650; AAA40515.1; -;  
 CC EMBL; BC033434; AAH33434.1; -;  
 CC PIR; I53066; I53066.  
 CC TRANSFAC; T01635; -;  
 CC MGD; MGI:98872; Twist1.  
 CC GO; GO:0005634; C:nucleus; NAS.  
 CC GO; GO:0003700; F:transcription factor activity; NAS.  
 CC GO; GO:0030154; P:cell differentiation; IMP.  
 CC GO; GO:0030326; P:limb morphogenesis; IMP.  
 CC GO; GO:0045596; P:negative regulation of cell differentiation; IDA.  
 CC GO; GO:0045843; P:negative regulation of myogenesis; IDA.  
 CC GO; GO:0001679; P:neuregulation; IMP.  
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
 CC InterPro; IPR001092; HLH\_basic.  
 CC Pfam; PF00010; HLH; 1.  
 CC SMART; SM00353; HLH; 1.  
 CC PROSITE; PS00038; HLH\_1; 1.  
 CC PROSITE; PS00888; HLH\_2; 1.  
 KW Differentiation; Developmental protein; Nuclear protein; DNA-binding;  
 KW Transcription regulation.  
 FT DOMAIN 80 102 GLY-RICH.  
 FT DOMAIN BIND 112 124 BASIC DOMAIN.

FT DOMAIN 125 164 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT VARIANT 36 36 A -> R (IN CDNA).  
 FT VARIANT 91 91 G -> P (IN CDNA).  
 SQ SEQUENCE 206 AA; 21198 MW; 618AD8B9B87C555 CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GCGAGGG 241  
 Db 84 GCGAGGG 90

RESULT 23  
 GLP1\_BRANA STANDARD; PRT; 207 AA.  
 AC P46271;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Germin-like protein 1 precursor.  
 GN GER1.  
 OS Brassica napus (Rape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPRAIN=cv. Samourai; Tissue=Seedling;  
 RX MEDLINE=98009982; PubMed=9349269;  
 RA Membré N., Berra A., Neutelings G., David A., David H., Staiger D.,  
 RA Saez Vaquez J., Raynal M., Delaeny M., Bernier P.;  
 RT "cDNA sequence, genomic organization and differential expression of  
 three Arabidopsis genes for germin/oxalate oxidase-like proteins.";   
 RL Plant Mol. Biol. 35:459-469(1997).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN PLANT DEFENSE. HAS PROBABLY NO  
 CC OXALATE OXIDASE ACTIVITY EVEN IF THE ACTIVE SITE IS CONSERVED.  
 CC -1- SUBUNIT: OLIGOMER (BELIEVED TO BE A PENTAMER BUT PROBABLY  
 CC HEXAMER) (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Apoplast (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE GERMIN FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; U21743; AA86365.1; -;  
 DR PIR; T07854; T07854.  
 DR HSSP; P45850; 1P12.  
 DR InterPro; IPR006045; Cupin.  
 DR InterPro; IPR007113; Cupin\_sup.  
 DR InterPro; IPR001929; Germin.  
 DR Pfam; PF00190; Cupin; 1.  
 DR PRINTS; PR00325; GERMIN.  
 DR PROSITE; PS00725; GERMIN; 1.  
 KW Apoplast; Cell wall; Signal; Glycoprotein; Manganese; Metal-binding.  
 FT SIGNAL 1 17  
 FT CHAIN 18 207  
 FT METAL 99 99 GERMIN-LIKE PROTEIN 1.  
 FT METAL 101 101 MANGANESE (BY SIMILARITY).  
 FT METAL 106 106 MANGANESE (BY SIMILARITY).  
 FT METAL 145 145 MANGANESE (BY SIMILARITY).  
 FT METAL 223 223 MANGANESE (BY SIMILARITY).  
 FT DISULFID 23 38 BY SIMILARITY.  
 FT CARBOHYD 58 58 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 207 AA; 21514 MW; B6A29465A73B956E CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TVKTLKG 10  
 Db 195 TVKTLKG 201

RESULT 24  
 RR3\_CHLVU STANDARD; PRT; 231 AA.  
 AC P56365;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chloroplast 30S ribosomal protein S3.  
 GN RPS3.  
 OS Chlorella vulgaris.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;  
 OC Chlorellaceae; Chlorella.  
 OX NCBI\_TaxID=3077;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IAM C-27 / Tamiya;  
 RX MEDLINE=97303241; PubMed=9159184;  
 RA Wakaesugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,  
 RA Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,  
 RA Inamura A., Yoshinaga K., Sugita M.;  
 RT "Complete nucleotide sequence of the chloroplast genome from the  
 green alga Chlorella vulgaris: the existence of genes possibly  
 involved in chloroplast division.";   
 RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).  
 CC -1- SUBUNIT: Part of the 30S ribosomal subunit.  
 CC -1- SUBCELLULAR LOCATION: Chloroplast.  
 CC -1- SIMILARITY: BELONGS TO THE S3P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -1- SIMILARITY: Contains 1 KH type-2 domain.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; AB001684; BAA58006.1; -;  
 DR PIR; T07358; T07358.  
 DR HAMAP; MF\_01309; -; 1.  
 DR InterPro; IPR004044; KH TYPE 2.  
 DR InterPro; IPR001351; Ribosomal\_S3.  
 DR InterPro; IPR005704; S3\_bact.  
 DR Pfam; PF00189; Ribosomal\_S3\_C; 1.  
 DR Pfam; PF00417; Ribosomal\_S3\_N; 1.  
 DR TIGRFAMs; TIGR01009; rpsC\_bact; 1.  
 DR PROSITE; PS00823; KH TYPE 2; 1.  
 DR PROSITE; PS00548; RIBOSOMAL\_S3; 1.  
 KW Ribosomal protein; RNA-binding; rRNA-binding; Chloroplast.  
 FT DOMAIN 39 123 KH TYPE-2  
 SQ SEQUENCE 231 AA; 26404 MW; D3EDE90A53551097 CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 VQALRA 171  
 Db 154 VQALRA 160

RESULT 25  
 CD8A\_FELCA STANDARD; PRT; 239 AA.  
 ID CD8A\_FELCA

P41688;  
01-NOV-1995 (Rel. 32, Created)  
01-NOV-1995 (Rel. 32, Last sequence update)  
15-SEP-2003 (Rel. 42, Last annotation update)  
T-cell surface glycoprotein CD8 alpha chain precursor.  
CD8A.  
GN Felis silvestris catus (Cat).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
OX NCBI\_TaxID=9685;  
RN [1]  
RP SEQUENCE FROM N.A.  
RQ TISSUE=Thymus;  
RX MEDLINE=94178799; PubMed=8132208;  
RA Pecoraro M., Kawaguchi Y., Miyazawa T., Norimine J., Maeda K.,  
RA Toyosaki T., Tohya Y., Kai C., Mikami T.;  
RT "Isolation, sequence and expression of a cDNA encoding the  
RT alpha-chain of the feline CD8.";  
RL Immunology 81:127-131(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RQ TISSUE=Thymus;  
RA Miyazawa T.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT  
CC WITH MHC CLASS I BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN  
CC THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO  
CC CLASS I MHC MOLECULES ALPHA-3 DOMAINS.  
CC -1- SUBUNIT: IN GENERAL HETERODIMER OF AN ALPHA AND A BETA CHAIN  
CC LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORM HOMODIMERS.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D16536; BAA03973.1; -.  
CC EMBL; AB000485; BAA19126.1; -.  
CC PIR; I46082; I46082.  
CC HSP; P01732; 1CD8.  
CC GO; GO:0042101; C:T-cell receptor complex; ISS.  
CC GO; GO:0015026; F:coreceptor activity; ISS.  
CC GO; GO:0042288; F:MHC class I protein binding activity; ISS.  
CC GO; GO:0005515; F:protein binding activity; ISS.  
CC GO; GO:0006955; P:immune response; ISS.  
CC GO; GO:0042110; P:T-cell activation; ISS.  
CC GO; GO:0007189; P:transmembrane receptor protein tyrosine kin. . .; ISS.  
CC InterPro; IPR007110; IG-like.  
CC InterPro; IPR003599; IG.  
CC Pfam; PF00047; Ig\_1.  
CC SMART; SM00409; IG\_1.  
CC PROSITE; PS0835; IG LIKE; 1.  
CC Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;  
KW Immune response; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 239  
FT T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA  
FT CHAIN.  
FT DOMAIN 22 188  
FT TRANSMEM 189 210  
FT POTENTIAL.  
FT DOMAIN 211 239  
FT CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 25 139  
FT IG-LIKE V-TYPE.  
FT DISULFID 46 119  
FT BY SIMILARITY.  
SQ SEQUENCE 239 AA; 26120 MW; F9A171820E8FF1BC CRC64;  
  
Query Match 1.9%; Score 7; DB 1; Length 239;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 APITTSQ 120  
DB 153 APITTSQ 159  
  
RESULT 26  
HMXU MUSDO STANDARD; PRT; 248 AA.  
AC Q254E1;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Homeotic Ultrabithorax protein (Fragment).  
GN UBX.  
OS Musca domestica (House fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Muscoidea;  
OC Muscidae; Musca.  
OX NCBI\_TaxID=7370;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wilde C.D., Akam M.;  
RT "Conserved sequence elements in the 5' region of the Ultrabithorax  
RT transcription unit.";  
RL EMBO J. 6:1393-1401(1987).  
CC -1- FUNCTION: Sequence-specific transcription factor which is part of  
CC a developmental regulatory system that provides cells with  
CC specific positional identities on the anterior-posterior axis.  
CC Binds the consensus region 5'-TTAAT[GT][GA]-3' (By similarity).  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X05178; CAA28813.1; -.  
CC InterPro; IPR001827; Antennapedia.  
CC InterPro; IPR001356; Homeobox.  
CC PROSITE; PS00032; ANTENNAPEDIA; 1.  
CC PROSITE; PS00027; HOMEBOX\_1; PARTIAL.  
KW Transcription regulation; Activator; Homeobox; DNA-binding;  
KW Developmental protein; Nuclear protein.  
FT NON\_TER 1 1  
FT DOMAIN 32 38 POLY-ALA.  
FT DOMAIN 108 123 POLY-GLY.  
FT DOMAIN 142 151 POLY-GLY.  
FT DOMAIN 224 227 POLY-ALA.  
FT SITE 241 246 ANTP-TYPE HEXAPEPTIDE.  
SQ SEQUENCE 248 AA; 23912 MW; AD7CE98CC39CCF2E CRC64;  
  
Query Match 1.9%; Score 7; DB 1; Length 248;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
DB 108 GGGAGGG 114  
  
RESULT 27  
GSHI HUMAN STANDARD; PRT; 264 AA.  
AC Q9H4S2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Homeobox protein GSH-1.

GN OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21588723; PubMed=11731616;  
 RA Mutsaers N., Iwasaki Y., Morishita M., Nomura A., Yamamori E.,  
 RT Yoshida M., Asai M., Ozaki N., Kambe F., Seo H., Oiso Y., Saito H.;  
 RA "Homeobox protein Gsh-1-dependent regulation of the rat GHRH gene  
 RT promoter."; Mol. Endocrinol. 15:2149-2156 (2001).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP Rates K.;  
 RA Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Probable transcription factor that binds to the DNA  
 CC sequence 5'-GC(TA)(AC)ATA(GA)-3'. Activates the transcription of  
 CC the GHRH gene. Plays an important role in pituitary development.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AB044157; BAB78692.1; -.  
 DR EMBL; AB044158; BAB78693.1; -.  
 DR EMBL; AL390738; CAC12721.1; -.  
 DR HSP; P14653; I872.  
 DR Genew; HGNC:20374; GSH1.  
 DR InterPro; IPR001356; Homeobox.  
 DR InterPro; IPR000047; HTH lambrapressr.  
 DR Pfam; PF00046; homeobox; 1.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR PRINTS; PR00031; HTHREPRESSR.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX 1; 1.  
 DR PROSITE; PS00071; HOMEBOX 2; 1.  
 KW Transcription regulation; Activator; Homeobox; DNA-binding;  
 KW Nuclear protein; Developmental protein.  
 FT DNA\_BIND 147 206 HOMEBOX.  
 FT DOMAIN 111 118 POLY-ALA.  
 FT DOMAIN 213 223 POLY-GLY.  
 SQ SEQUENCE 264 AA; 27882 MW; 25F4C4336E270C00 CRC64;  
 Query Match 1.9%; Score 7; DB 1; Length 264;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 235 GGGAGGG 241  
 DB 216 GGGAGGG 222  
 |||||  
 RESULT 28  
 CANS\_RABIT  
 ID\_CANS\_RABIT STANDARD; PRT; 266 AA.  
 AC P06813;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)  
 DE (Calcium-activated neutral proteinase) (CANP).  
 GN CAPNS1 OR CAPNA4.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86250503; PubMed=3013892;  
 RA Emori Y., Kawasaki H., Imajoh S., Kawashima S., Suzuki K.;  
 RT "Isolation and sequence analysis of cDNA clones for the small subunit  
 RT of rabbit calcium-dependent protease."; J. Biol. Chem. 261:9472-9476(1986).  
 RL [2]  
 RN CALCIUM-BINDING DATA.  
 RX MEDLINE=87279982; PubMed=3038955;  
 RA Minami Y., Emori Y., Kawasaki H., Suzuki K.;  
 RT "B-P hand structure-domain of calcium-activated neutral protease  
 RT (CANP) can bind Ca<sup>2+</sup> ions."; J. Biochem. 101:889-895(1987).  
 RL [2]  
 CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which  
 CC catalyze limited proteolysis of substrates involved in  
 CC cytoskeletal remodeling and signal transduction.  
 CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small  
 CC (regulatory) subunit.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma  
 CC membrane upon Ca<sup>2+</sup> binding (By similarity).  
 CC -1- PTM: The N-terminus is blocked.  
 CC -1- SIMILARITY: Contains 4 EF-hand calcium-binding domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M13364; AAA81565.1; -.  
 DR PIR; A24816; CIRBL.  
 DR HSP; P04632; IKFU.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00036; efhand; 3.  
 DR SMART; SM00054; EFh; 3.  
 DR PROSITE; PS00018; EF\_HAND; 2.  
 KW Calcium-binding; Repeat.  
 FT DOMAIN 1 64 GLY-RICH (HYDROPHOBIC).  
 FT DOMAIN 94 266 CALCIUM-BINDING.  
 FT CA\_BIND 150 161 EF-HAND 1 (POTENTIAL).  
 FT CA\_BIND 180 191 EF-HAND 2 (POTENTIAL).  
 FT DOMAIN 215 226 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).  
 FT DOMAIN 245 258 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).  
 FT DOMAIN 10 25 POLY-GLY.  
 FT DOMAIN 34 54 POLY-GLY.  
 FT DOMAIN 76 81 POLY-PRO.  
 SQ SEQUENCE 266 AA; 28239 MW; 1D7FE31989F70B03 CRC64;  
 Query Match 1.9%; Score 7; DB 1; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 235 GGGAGGG 241  
 DB 47 GGGAGGG 53  
 |||||  
 RESULT 29  
 EP34\_HCMVA  
 ID EP34\_HCMVA STANDARD; PRT; 268 AA.  
 AC P16768;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Early phosphoprotein P34.  
 GN UL112.  
 OS Human cytomegalovirus (strain AD169).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

```

CC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA "Reddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RA "Analysis of the protein-coding content of the sequence of human
RA cytomegalovirus strain AD169."
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Wang S.K., Duh C.Y.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-6 AND
CC HSV-7 U79 AND HCMV UL112 (P34).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17403; CAA35315.1; ALT_TERM.
DR EMBL; U57433; AAB53252.1; -.
DR PIR; S09860; S09860.
DR InterPro; IPR004138; U79_P34.
DR Pfam; PF03064; U79_P34; 1.
KW Early protein; Phosphorylation.
FT DOMAIN 148 153 POLY-GLY.
FT DOMAIN 197 200 POLY-SER.
FT DOMAIN 202 220 POLY-GLY.
SQ SEQUENCE 268 AA; 28142 MW; BD5906F4EAB28FC9 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241
Db 206 GGGAGGG 212
|||||

RESULT 30
CDK4_PIG STANDARD; PRT; 303 AA.
AC P79432;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Cell division protein kinase 4 (EC 2.7.1.-) (Cyclin-dependent kinase
DE 4).
GN CDK4.
OS Sus scrofa (Pig)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Shibuya H., Renshaw F.G., Bouchard G.F., Nonneman D., Johnson G.S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY INVOLVED IN THE CONTROL OF THE CELL CYCLE.
CC -!- ENZYME REGULATION: PHOSPHORYLATION AT T-172 IS NECESSARY FOR
CC ENZYMACTIC ACTIVITY (BY SIMILARITY).
CC -!- SUBUNIT: FORMS A STABLE COMPLEX WITH D-TYPE G1 CYCLINS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CDC2/CDKX SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U68478; AAB39636.1; -.
DR EMBL; U68479; AAB39637.1; -.
DR HSSP; Q00534; 1B18.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00669; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Cell cycle; Cell division; Phosphorylation.
FT DOMAIN 6 295
FT NP_BIND 12 20 ATP (BY SIMILARITY).
FT BINDING 35 35 ATP (BY SIMILARITY).
FT ACT_SITE 140 140 BY SIMILARITY.
FT MOD_RES 172 172 PHOSPHORYLATION (BY CAX) (BY SIMILARITY).
SQ SEQUENCE 303 AA; 33690 MW; D77C5630B25EF20B CRC64;

Query Match 1.9%; Score 7; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241
Db 42 GGGAGGG 48
|||||

RESULT 31
RALLY_HUMAN STANDARD; PRT; 306 AA.
ID RALLY_HUMAN
AC Q9UKM9; Q14621; Q9BQX6; Q9UJE3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA-binding protein Raly (hnRNP associated with lethal yellow homolog)
DE (Autoantigen p542).
GN RALY OR P452.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Testis;
RX MEDLINE=99431566; PubMed=10500250;
RA Kirebtukova I., Kuklin A., Woychik R.P., Michaud E.J.;
RT "Alternative processing of the human and mouse raly genes.";
RL Biochim. Biophys. Acta 1447:107-112(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Vaughan J.H.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Scriver K.F., Bates K.N., Bead L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

```



RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.B., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lehesvaisto M.H., Leveraha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McComachie L.J., McKay K., McMurray A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whithead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871 (2001).  
 RN [4]  
 RP SEQUENCE OF 85-306 FROM N.A. (ISOFORM 1), AND AUTOIMMUNE DISEASE.  
 RC TISSUE=Lymphocytes;  
 RX MEDLINE=98018738; PubMed=9376072;  
 RA Rhodes G.H., Valbracht J.R., Nguyen M.-D., Vaughan J.H.;  
 RT "The p542 gene encodes an autoantigen that cross-reacts with EBNA-1 of  
 RT the Epstein Barr virus and which may be a heterogeneous nuclear  
 RT ribonucleoprotein.";  
 RL J. Autoimmun. 10:447-454 (1997).  
 RN [5]  
 RP SEQUENCE OF 227-253, AND DETERMINATION OF AUTOANTIGENIC EPITOPE.  
 RX MEDLINE=95190029; PubMed=7533789;  
 RA Vaughan J.H., Valbracht J.R., Nguyen M.-D., Handley H.H., Smith R.S.,  
 RA Patrick K., Rhodes G.H.;  
 RT "Epstein-Barr virus-induced autoimmune responses. I. Immunoglobulin M  
 RT autoantibodies to proteins mimicking and not mimicking Epstein-Barr  
 RT virus nuclear antigen-1.";  
 RL J. Clin. Invest. 95:1306-1315 (1995).  
 CC -!- FUNCTION: Probable-RNA binding protein. Could be a heterogeneous  
 CC nuclear ribonucleoprotein (hnRNP).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=2;  
 CC IsoId=Q9UKM9-1; Sequence=Displayed;  
 CC Name=1;  
 CC IsoId=Q9UKM9-2; Sequence=VSP\_005804;  
 CC -!- TISSUE SPECIFICITY: Expressed in heart, brain, lung, liver,  
 CC skeletal muscle, kidney and pancreas. Weakly expressed in  
 CC placenta.  
 CC -!- DISEASE: Autoantigen found in infectious mononucleosis caused by  
 CC Epstein-Barr virus. An epitope recognized by B-cells, which cross-  
 CC react with the BKF1 protein (EBNA-1 nuclear protein) of Epstein-  
 CC Barr virus has been identified.  
 CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.  
 CC -!- CAUTION: Ref.4 (CAC29371) sequence differs from that shown due to  
 CC erroneous gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF148457; AAF04487.1; -;  
 CC EMBL; L38696; AAC28898.1; -;  
 CC EMBL; AL031668; CAC29371.1; ALT\_SEQ.  
 CC EMBL; AL031668; CAB43742.1; -;  
 CC Genew; HGNC:15921; RALY.  
 CC InterPro; IPR000504; RNA\_rec\_mot.  
 CC Pfam; PF00076; rrm; 1.  
 CC SMART; SM00360; RRM; 1.  
 CC PROSITE; PS50102; RRM; 1.

DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 KW Ribonucleoprotein; RNA-binding; Nuclear protein; Antigen;  
 KW Alternative splicing; Polymorphism.  
 FT DOMAIN 21 92 RNA-BINDING (RRM).  
 FT DOMAIN 227 253 EPITOPE (RECOGNIZED BY BKRFL ANTIBODIES).  
 FT DOMAIN 225 251 POLY-GLY.  
 FT VARSPIC 110 125 Missing (in isoform 1).  
 FT VARIANT 215 215 /FTID=VSP\_005804.  
 FT VARIANT 251 251 /FTID=VAR\_015223.  
 FT VARIANT 251 251 /FTID=VAR\_015224.  
 FT CONFLICT 214 215 EQ -> DE (IN REF. 2).  
 FT CONFLICT 230 230 A -> AS (IN REF. 2).  
 SQ SEQUENCE 306 AA; 32463 MW; 7F4376D3BD8E4728 CRC64;  
 Query Match 1.9%; Score 7; DB 1; Length 306;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 235 GGGAGGG 241  
 Db 227 GGGAGGG 233  
 RESULT 32  
 ID DDL\_XYLFA STANDARD; PRT; 320 AA.  
 AC Q9P79;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE D-alanine-D-alanine ligase (EC 6.3.2.4) (D-alanylalanine synthetase)  
 DE (D-Ala-D-Ala ligase).  
 GN DDL OR XF0799.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9a5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferto J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohnsbeil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lenos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Mend C.F.N., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pequerio J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.B. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Souza J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";  
 RL Nature 406:151-159 (2000).  
 CC -!- FUNCTION: Cell wall formation (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + 2 D-alanine = ADP + phosphate + D-





RA Ananatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Agayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu C., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmei B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Murny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-f., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195 (2000).  
RN [3]  
RP REVISIONS.  
RC STRAIN=Berkeley;  
RX MEDLINE=22426059; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review."  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley; TISSUE=Testis;  
RX MEDLINE=22426066; PubMed=12537569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
RA Rubin G.M., Celisner S.E.;  
RT "A *Drosophila* full-length cDNA resource."  
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
RN [5]  
RP SEQUENCE OF 24-289 FROM N.A.  
RC STRAIN=WS1, WS9, WS12, WS16, WS17, WS19, WS26, WS47, ZIM5, ZIM10,  
RC ZIM22, ZIM24, ZIM30, ZIM32, and ZIM35;  
RX MEDLINE=20556153; PubMed=1102381;  
RA Begun D.J., Whitley P., Todd B.L., Waldrip-Dail H.M., Clark A.G.;  
RT "Molecular population genetics of male accessory gland proteins in  
RT *Drosophila*."  
RL Genetics 156:1879-1888(2000).  
CC -!- FUNCTION: Responsible for physiological and behavioral changes in  
CC mated female flies.  
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).  
CC -!- ALTERNATIVE PRODUCTS:

CC Event-Alternative Initiation;  
CC Comment=2 isoforms, Long (shown here) and Short, are produced by  
CC alternative initiation;  
CC -!- TISSUE SPECIFICITY: Seminal fluid.  
CC -!- CAUTION: Ref.3 and Ref.4 sequences differ from that shown due to  
CC frameshifts. These may be a natural frameshifts and this protein  
CC could be a pseudogene.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U85764; AAB96388.1; -  
CC EMBL; U85764; AAB96389.1; -  
CC EMBL; U90948; AAB96395.1; -  
CC EMBL; AE003630; AAF53055.2; ALT\_FRAME.  
CC EMBL; AY089245; AAL89983.1; ALT\_FRAME.  
CC EMBL; AY010544; AAG32576.1; -  
CC EMBL; AY010545; AAG32577.2; -  
CC EMBL; AY010546; AAG32578.1; -  
CC EMBL; AY010547; AAG32579.1; -  
CC EMBL; AY010548; AAG32580.1; -  
CC EMBL; AY010549; AAG32581.2; -  
CC EMBL; AY010550; AAG32582.1; -  
CC EMBL; AY010551; AAG32583.1; -  
CC EMBL; AY010552; AAG32584.1; -  
CC EMBL; AY010553; AAG32585.1; -  
CC EMBL; AY010554; AAG32586.1; -  
CC EMBL; AY010555; AAG32587.1; -  
CC EMBL; AY010556; AAG32588.1; -  
CC EMBL; AY010557; AAG32589.1; -  
CC EMBL; AY010558; AAG32590.1; -  
CC EMBL; AY010559; AAG32591.1; -  
CC FlyBase; FBgn0023415; Acp32CD.  
CC GO; GO:0005576; C:extracellular; IE.  
CC GO; GO:0005180; F:peptide hormone; NAS.  
CC GO; GO:0045434; P:negative regulation of female receptivity, . . .; NAS.  
CC GO; GO:0045434; P:negative regulation of female receptivity, . . .; NAS.  
KW Alternative Initiation; Signal; Behavior; Polymorphism.  
FT SIGNAL 1 21  
FT CHAIN 22 329  
FT CHAIN 89 329  
FT CHAIN 89 329  
FT INIT MET 89 89  
FT DOMAIN 100 182  
FT DOMAIN 200 326  
FT VARIANT 45 45  
FT VARIANT 52 52  
FT VARIANT 63 63  
FT VARIANT 69 69  
FT VARIANT 273 275  
FT VARIANT 275 275  
FT SEQUENCE 329 AA; 35557 MW; 5655CCCA73759CFA CRC64;  
SQ  
Query Match 1.9%; Score 7; DB 1; Length 329;  
Best Local Similarity 100.0%; Pred.No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 235 GGGAGGG 241  
DB 110 GGGAGGG 116

```
RESULT 35
TAL_MOUSE STANDARD; PRT; 329 AA.
ID TAL_MOUSE
AC P22091;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE T-cell acute lymphocytic leukemia-1 protein homolog (TAL-1 protein)
DE (Stem cell protein).
GN TAL1 OR TAL-1 OR SCL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Bone marrow macrophage;
RA MEDLINE=91126096; PubMed=1704135;
RA Begley C.G., Visvader J., Green A.R., Aplan P.D., Metcalf D.,
RA Kirsch I.R., Gough N.M.;
RT "Molecular cloning and chromosomal localization of the murine homolog
RT of the human helix-loop-helix gene SCL."
RL Proc. Natl. Acad. Sci. U.S.A. 88:869-873 (1991).
CC -1- FUNCTION: IMPLICATED IN THE GENESIS OF HEMOPOIETIC MALIGNANCIES.
CC IT MAY PLAY AN IMPORTANT ROLE IN HEMOPOIETIC DIFFERENTIATION.
CC SERVES AS A POSITIVE REGULATOR OF ERYTHROID DIFFERENTIATION.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. FORMS HETERODIMERS WITH E12 AND E47. BINDS TO THE
CC LIM DOMAIN CONTAINING PROTEIN RHOMBOTIN-2 AND TO DRG1.
CC -1- TISSUE SPECIFICITY: ERYTHROID AND MYELOID CELLS.
CC -1- DOMAIN: THE HELIX-LOOP-HELIX DOMAIN IS NECESSARY AND SUFFICIENT
CC FOR THE INTERACTION WITH DRG1.
CC -1- PTM: PHOSPHORYLATED ON SERINE RESIDUES (BY SIMILARITY).
CC -1- DISEASE: INVOLVED IN CHROMOSOMAL TRANSLOCATION IN LEUKEMIC
CC STEM-CELLS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59764; AAA40097.1; -
DR EMBL; U01530; AAA86937.1; -
DR PIR; A37864; A37864.
DR TRANSFAC; T01799; -
DR MGD; MGI:98480; Tal1.
DR GO; GO:0030218; P:erythrocyte differentiation; IDA.
DR GO; GO:0030097; P:hemopoiesis; IDA.
DR InterPro; IPR001092; HLH_bas.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS50888; HLH_2; 1.
KW DNA-binding; Transcription regulation; Differentiation;
KW Proto-oncogene; Chromosomal translocation; Phosphorylation.
FT DNA BIND 187 199 BASIC DOMAIN.
FT DOMAIN 200 240 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 263 272 POLY-GLY.
SQ SEQUENCE 329 AA; 34278 MW; 189480B6993CB371 CRC64;
Query Match 1.9%; Score 7; DB 1; Length 329;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 235 GGGAGGG 241
DB 266 GGGAGGG 272
```

```
RESULT 36
SIX3_HUMAN STANDARD; PRT; 332 AA.
ID SIX3_HUMAN
AC G95343;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Homeobox protein SIX3 (Sine oculis homeobox homolog 3).
GN SIX3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99107815; PubMed=9889003;
RA Granadino B., Gallardo M.E., Lopez-Rios J., Sanz R., Ramos C.,
RA Ayuso C., Bovolenta P., Rodriguez de Cordoba S.;
RT "Genomic cloning, structure, expression pattern, and chromosomal
RT location of the human SIX3 gene."
RL Genomics 55:100-105 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99377859; PubMed=10454822;
RA Leppert G.S., Yang J.-M., Sundin O.H.;
RT "Sequence and location of SIX3, a homeobox gene expressed in the human
RT eye."
RL Ophthalmic Genet. 20:1-15 (1999).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Retina;
RA Clark B.J., Hanson I.M., Brown A.G., Ferrier R.K., Prosser J.,
RA van Heyningen V.;
RT "SIX3, a member of the Sine oculis/Six family of transcription
RT factors, is expressed in the developing and adult human eye."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP VARIANTS HPE2 VAL-226; ALA-250 AND PRO-257.
RX MEDLINE=99295940; PubMed=10369266;
RA Wallis D.E., Roessler E., Hehr U., Nanni L., Wiltshire T.,
RA Richter-Costa A., Gillespie-Kaesbach G., Zackai E.H., Rommens J.,
RA Muenke M.;
RT "Mutations in the homeodomain of the human SIX3 gene cause
RT holoprosencephaly."
RL Nat. Genet. 22:196-198 (1999).
CC -1- FUNCTION: MAY BE INVOLVED IN VISUAL SYSTEM DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: Defects in SIX3 are the cause of holoprosencephaly type 2
CC (HPE2) [MIM:157170]. HPE2 is a common, severe malformation of the
CC brain that involves separation of the central nervous system into
CC left and right halves.
CC -1- SIMILARITY: BELONGS TO THE SIX3/SINE OCULIS HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF092047; AAD11939.1; -
DR EMBL; AF049339; AAD15753.1; -
DR EMBL; AF083891; AAD51091.1; -
DR EMBL; AJ012611; CAB42539.1; -
DR HSSP; P41778; 1DU6.
DR TRANSFAC; T03282; -
DR Genew; HGNC:10889; SIX3.
DR MIM; 603714; -
DR MIM; 157170; -
DR GO; GO:0007420; P:brain development; TAS.
```

```

DR GO: 0007601; P: vision; TAS.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR000047; HTH_lambdarepressor.
DR InterPro: IPR007105; SIX.
DR InterPro: IPR007106; SIX_SINE_homeo.
DR Pfam: PF000046; homeobox_1.
DR PRINTS: PR00031; HTHRPRSSR.
DR PRODOM: PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
DR PROSITE; PS00071; HOMEBOX 2; 1.
DR Developmental protein; Homeobox; DNA-binding; Nuclear protein;
KW Disease mutation; Holoprosencephaly.
FT DOMAIN 33 69 POLY-GLY.
FT DOMAIN 206 265 HOMEBOX.
FT DOMAIN 263 266 POLY-ALA.
FT VARIANT 226 226 L -> V (in HPE2).
FT VARIANT 250 250 V -> A (in HPE2).
FT VARIANT 257 257 R -> P (in HPE2).
FT VARIANT 257 257 /FTID=VAR_003772.
FT VARIANT 257 257 /FTID=VAR_003773.
SQ SEQUENCE 332 AA; 35486 MW; 21EA07F6A2DD978F CRC64;

Query Match 1.9%; Score 7; DB 1; Length 332;
Best Local Similarity 100.0%; Pred.No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241
DB 41 GGGAGGG 47

RESULT 37
SIX3 MOUSE STANDARD; PRT; 333 AA.
AC Q62233; P70176; P70177;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein SIX3 (Sine oculis homeobox homolog 3).
GN SIX3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Embryonic brain;
RX MEDLINE=96125147; PubMed=8575305;
RA Oliver G., Mailhos A., Wehr R., Copeland N.G., Jenkins N.A.,
RA Gruss P.;
RT "Six3, a murine homologue of the sine oculis gene, demarcates the
RT most anterior border of the developing neural plate and is expressed
RT during eye development.";
RL Development 121:4045-4055(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=96409319; PubMed=8814301;
RA Kawakami K., Ohno H., Takizawa T., Saito T.;
RT "Identification and expression of Six family genes in mouse retina.";
RL FEBS Lett. 393:259-263(1996).
CC -1- FUNCTION: May be involved in visual system development.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Six3B;
CC IsoId=Q62233-1; Sequence=Displayed;
CC Name=Six3A;
CC IsoId=Q62233-2; Sequence=VSP_002291, VSP_002292;
CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT E6.5 OF EMBRYO DEVELOPMENT
CC AROUND THE ANTERIOR BORDER. AT E8.5, EXPRESSION IS FOUND OVER THE

```

```

CC ANTERIOR NEURAL PLATE. AT E9.5, IN THE DIENCEPHALIC PART OF THE
CC VENTRAL FOREBRAIN, OPTIC VESICLES, OLFACTORY PLACODES AND RATHKE'S
CC POUCH. IN LATER STAGES, PRESENT IN HYPOTHALAMUS, EYES AND
CC PITUITARY.
CC -1- SIMILARITY: BELONGS TO THE SIX/SINE OCULIS HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X90871; CAAG2379.1; ALT_INIT.
CC EMBL; D83144; BAA11822.1; -.
CC EMBL; D83145; BAA11823.1; -.
CC HSSP; P41778; 1DU6.
CC TRANSFAC; T03263; -.
CC TRANSFAC; T03270; -.
CC MGD; MGI:102764; Six3.
CC GO: GO:0005634; C:nucleus; IDA.
CC GO: GO:0005515; F:protein binding activity; IPI.
CC GO: GO:0014564; F:transcriptional repressor activity; IDA.
CC GO: GO:0001654; P:eye morphogenesis; IDA.
CC GO: GO:0016481; P:negative regulation of transcription; IDA.
CC InterPro: IPR001356; Homeobox.
CC InterPro: IPR007105; SIX.
CC InterPro: IPR007106; SIX_SINE_homeo.
CC Pfam; PF00046; homeobox; 1.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
CC PROSITE; PS00071; HOMEBOX 2; 1.
KW Developmental protein; Homeobox; DNA-binding; Nuclear protein;
KW Alternative splicing.
FT DOMAIN 33 70 GLY-RICH.
FT DNA_BIND 207 266 HOMEBOX.
FT DOMAIN 264 267 POLY-ALA.
FT VARSPLIC 271 286 LQQAIGPSGMRSLAE -> SVAGTAARPPQA (in
FT isoform SIX3A).
FT /FTID=VSP_002291.
FT Missing (in isoform SIX3A).
FT /FTID=VSP_002292.
FT G -> GG (IN REF. 1).
FT VA -> WP (IN REF. 1).
FT CONFLICT 118 119 PGMRLAEPGCPTHGSAESPSTAASPTTSVSLTERADTG
FT CONFLICT 278 333 TSILSVTSSDSDCDV -> ERDALPGARLPHARLSRVTH
FT GGGPDHQCVCQPDGARGHRHFDPLGNLRLGM (IN REF.
FT 1).
SQ SEQUENCE 333 AA; 35592 MW; 1AD7D3C4388043B9 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 333;
Best Local Similarity 100.0%; Pred.No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241
DB 42 GGGAGGG 48

RESULT 38
SYEA_CLOAB STANDARD; PRT; 339 AA.
AC Q97GK9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phenylalanine-tRNA synthetase alpha chain (EC 6.1.1.20)
DE (Phenylalanine--tRNA ligase alpha chain) (PhRS).
GN PHRS OR CAC2357.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

```

CC Clostridium.  
 OX NCBI\_TaxID=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 RA MEDLINE=21359325; PubMed=11466286;  
 RX Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum.";  
 RL J. Bacteriol. 183:4823-4838 (2001).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanyl-tRNA(Phe) = AMP +  
 CC di-phosphate + L-phenylalanyl-tRNA(Phe).  
 CC -1- COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).  
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
 CC Phe-tRNA synthetase alpha chain subfamily 1.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; AE007736; AAK80313.1; -  
 DR PIR; F97190; F97190.  
 DR HAMAP; MF\_00281; -; 1.  
 DR InterPro; IPR004188; Phe\_tRNA\_synt\_N.  
 DR InterPro; IPR004529; PheS.  
 DR InterPro; IPR002319; tRNA-synt 2d.  
 DR InterPro; IPR006195; tRNA-synt N; 1.  
 DR Pfam; PF02912; Phe\_tRNA-synt\_N; 1.  
 DR Pfam; PF01409; tRNA-synt 2d; 1.  
 DR TIGRFAMs; TIGR00468; pheS; 1.  
 DR PROSITE; PS50862; AA tRNA\_LIGASE II; 1.  
 DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Metal-binding; Magnesium; Complete proteome.  
 FT METAL 254 254 MAGNESIUM (BY SIMILARITY).  
 SQ SEQUENCE 339 AA; 38508 MW; 229E76449BE10918 CRC64;  
 -----  
 Query Match 1.9%; Score 7; DB 1; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 QY 27 KKNIEEI 33  
 Db 22 KKNIEEI 28  
 |||||  
 |||||  
 -----  
 RESULT 39  
 ID ROAL SCHAM STANDARD; PRT; 342 AA.  
 AC P21522;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Heterogeneous nuclear ribonucleoprotein A1, A2/B1 homolog.  
 GN HNRNP.  
 OS Schistocerca americana (American grasshopper).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Acridomorpha;  
 OC Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridoidea;  
 OC Acridoidea; Acrididae; Cyrtacanthacridinae; Schistocerca.  
 OX NCBI\_TaxID=7009;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91195067; PubMed=1901646;  
 RA Ball E.E., Rehm E.J., Goodman C.S.;

RT "Cloning of a grasshopper cDNA coding for a protein homologous to the  
 AL, A2/B1 proteins of mammalian hnRNP.";  
 RL Nucleic Acids Res. 19:397-397(1991).  
 CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF RIBONUCLEOSOMES.  
 CC -1- DEVELOPMENTAL STAGE: ABUNDANT IN EMBRYOS YOUNGER THAN 50% OF  
 CC EMBRYONIC DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE A/B GROUP OF HNRNP, WHICH ARE BASIC AND  
 CC GLY-RICH PROTEINS.  
 CC -1- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; X54670; CAA38481.1; -  
 DR PIR; S14432; S14432.  
 DR HGSP; P09651; IUP1.  
 DR InterPro; IPR005054; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 2.  
 DR SMART; SM00360; RRM; 2.  
 DR PROSITE; PS0102; RRM; 2.  
 DR PROSITE; PS00030; RRM\_RNP 1; 2.  
 KW Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.  
 FT DOMAIN 17 93 RNA-BINDING (RRM) 1.  
 FT DOMAIN 108 185 RNA-BINDING (RRM) 2.  
 FT DOMAIN 187 339 GLY-RICH.  
 SQ SEQUENCE 342 AA; 35404 MW; F8426A21C8125DD8 CRC64;  
 -----  
 Query Match 1.9%; Score 7; DB 1; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 QY 235 GGGAGGG 241  
 Db 202 GGGAGGG 208  
 |||||  
 |||||  
 -----  
 RESULT 40  
 ID BET3 MESAU STANDARD; PRT; 367 AA.  
 AC O09029;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE BET3 protein.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 OX NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96140430; PubMed=9552091;  
 RA Peyton M., Stellrecht C.M.M., Naya F.J., Huang H.-P., Samora P.J.,  
 RA Tsai M.-J.;  
 RT "BET3, a novel helix-loop-helix protein, can act as a negative  
 RT regulator of BET2 and MyoD-responsive genes.";  
 RL Mol. Cell. Biol. 16:626-633 (1996).  
 CC -1- FUNCTION: INHIBITS DNA BINDING OF TCF3 (B47) HOMODIMERS AND TCF3  
 CC (B47) / NEUROD1 HETERODIMERS AND ACTS AS A STRONG REPRESSOR OF  
 CC NEUROD1 AND MYOD-RESPONSIVE GENES, PROBABLY BY HETERODIMERIZATION  
 CC WITH CLASS A BASIC HELIX-LOOP-HELIX FACTORS. DESPITE THE PRESENCE  
 CC OF AN INTACT BASIC DOMAIN, DOES NOT BIND TO DNA.  
 CC -1- SUBUNIT: HETERODIMER WITH OTHER BHLH PROTEINS, LIKE TCF3 (B47).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- TISSUE SPECIFICITY: KIDNEY, LUNG, BRAIN AND PANCREAS (INSULINOMA).  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.  
 -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; S80870; AAB50691.1; -;  
 CC TRANSFAC; T01674; -;  
 CC InterPro; IPR001092; HLH\_basic.  
 CC Pfam; PF00010; HLH; 1;  
 CC SMART; SM00353; HLH; 1;  
 CC PROSITE; PS00038; HLH 1; 1;  
 CC PROSITE; PS00888; HLH 2; 1;  
 CC Nuclear protein; Transcription regulation; Repressor.  
 KW DOMAIN 11 14 POLY-ALA.  
 FT DOMAIN 58 62 POLY-SER.  
 FT DOMAIN 83 99 POLY-GLY.  
 FT DOMAIN 174 179 POLY-GLY.  
 FT DOMAIN 204 217 POLY-GLY.  
 FT DNA BIND 229 240 BASIC DOMAIN.  
 FT DOMAIN 241 282 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 311 319 POLY-ALA.  
 SQ SEQUENCE 367 AA; 35905 MW; 6CAB9AFF96B58F77 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 367;

Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241

DB 85 GGGAGGG 91

# RESULT 41

NLPD\_SALTI  
 ID NLPD\_SALTI STANDARD; PRT; 373 AA.  
 AC Q56131;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Lipoprotein nlpd precursor.  
 GN NLPD OR STV3050 OR T2826.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrall B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ty2 / ATCC 700931;  
 RX MEDLINE=22531367; PubMed=12644504;  
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
 RT and CT18.";  
 RL J. Bacteriol. 185:2330-2337(2003).  
 RN [3]

RP SEQUENCE OF 255-373 FROM N.A.  
 RC STRAIN=Ty21a;  
 RX MEDLINE=95220644; PubMed=7705608;  
 RA Robbe-Saule V., Coyneault C., Norel F.;  
 RT "The live oral typhoid vaccine Ty21a is a rps8 mutant and is  
 RT susceptible to various environmental stresses.";  
 RL FEMS Microbiol. Lett. 126:171-176(1995).  
 RN [4]  
 RP SEQUENCE OF 269-373 FROM N.A.  
 RC STRAIN=Ty2 / ATCC 700931;  
 RX MEDLINE=99118315; PubMed=9919662;  
 RA Robbe-Saule V., Norel F.;  
 RT "The rps8 mutant allele of Salmonella typhi Ty2 is identical to that  
 RT of the live typhoid vaccine Ty21a.";  
 RL FEMS Microbiol. Lett. 170:141-143(1999).  
 CC -|- FUNCTION: MAY BE INVOLVED IN STATIONARY-PHASE SURVIVAL.  
 CC -|- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid  
 CC anchor (Potential).  
 CC -|- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AL627276; CAD06031.1; -;  
 EMBL; AF016843; AA070383.1; -;  
 EMBL; X81641; CAA57297.1; -;  
 EMBL; Y17610; CAA76806.1; -;  
 MEROPS; M37.UFW; -;  
 InterPro; IPR002482; LysM.  
 InterPro; IPR002886; Peptidase\_M37.  
 Pfam; PF01476; LysM; 1.  
 Pfam; PF01551; Peptidase\_M37; 1.  
 SMART; SM00257; LysM; 1.  
 PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.  
 KW Inner membrane; Lipoprotein; Repeat; Signal; Complete proteome.  
 FT SIGNAL 1 25 BY SIMILARITY.  
 FT CHAIN 26 373 LIPOPROTEIN NLPD.  
 FT LIPID 26 26 N-ACYL DIGLYCERIDE (BY SIMILARITY).  
 FT DOMAIN 199 246 4 X 7 AA APPROXIMATE REPEATS.  
 FT REPEAT 199 205 1.  
 FT REPEAT 221 227 2.  
 FT REPEAT 233 239 3.  
 FT REPEAT 240 246 4.  
 SQ SEQUENCE 373 AA; 39183 MW; DD075BB440098FF7 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 373;

Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 GSTGTSS 87

DB 344 GSTGTSS 350

# RESULT 42

GLK1\_TRIVA  
 ID GLK1\_TRIVA STANDARD; PRT; 375 AA.  
 AC Q9GTW9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glucokinase 1 (EC 2.7.1.2) (Glucose kinase 1) (Hexokinase 1).  
 GN GK1.  
 OS Trichomonas vaginalis.  
 OC Eukaryota; Parabasalides; Trichomonadida; Trichomonadidae;  
 OC Trichomonadinae; Trichomonas.  
 OX NCBI\_TaxID=5722;  
 RN [1]

RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
RC STRAIN=ATCC 30001 / NIH-C1;  
RX MEDLINE=21147937; PubMed=11250082;  
RA Wu G., Henze K., Mueller M.;  
RT "Evolutionary relationships of the glucokinase from the  
RL amitochondriate protist, Trichomonas vaginalis.";  
CC Gene 264:265-271(2001).  
CC -|- CATALYTIC ACTIVITY: ATP + D-glucose = ADP + D-glucose 6-phosphate.  
CC -|- SUBUNIT: Monomer.  
CC -|- PTM: The N-terminus is blocked.  
CC -|- SIMILARITY: BELONGS TO THE BACTERIAL GLUCOKINASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF248652; AAG17616.1; -;  
DR Transferase; Kinase; Glycolysis; ATP-binding.  
KW NP\_BIND 25 30 ATP (POTENTIAL)  
FT SEQUENCE 375 AA; 41585 MW; 229F76B24637BE21 CRC64;  
SQ  
  
Query Match 1.9%; Score 7; DB 1; Length 375;  
Best Local Similarity 100.0%; Pred.No. 48;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 205 TDRAPTQ 211  
Db 155 TDRAPTQ 161  
|||||  
155 TDRAPTQ 161  
  
RESULT 43  
DNAJ LISIN  
ID DNAJ LISIN STANDARD; PRT; 376 AA.  
AC Q92BN9;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Chaperone protein dnaJ.  
GN DNAJ OR L1N1509.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=CLIP 11262 / Serovar 6a;  
RC MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel N., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunet F., Kurapkat G.,  
RA Madueno E., Maitournan A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001).  
CC -|- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,  
CC THE APASE ACTIVITY OF DNAK (BY SIMILARITY).  
CC -|- CAPCITOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -|- SIMILARITY: BELONGS TO THE DNAJ FAMILY.  
CC -|- SIMILARITY: Contains 1 J domain.  
CC -|- SIMILARITY: Contains 1 CR domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL596168; CAC96740.1; -;  
DR PIR; AD1621; AD1621.  
DR L1stiliat; L1N01509; -;  
DR InterPro; IPR002939; DnaJ C.  
DR InterPro; IPR001305; DnaJ\_CXXCXGKG.  
DR InterPro; IPR001623; DnaJ\_N.  
DR InterPro; IPR003095; Hsp\_DnaJ.  
DR Pfam; PF00226; DnaJ; 1.  
DR Pfam; PF01556; DnaJ C; 1.  
DR Pfam; PF00684; DnaJ\_CXXCXGKG; 1.  
DR PRINTS; PR00625; DNAJPROTEIN.  
DR SMART; SM00271; DnaJ; 1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PS00767; DnaJ\_2; 1.  
DR PROSITE; PS00637; DnaJ\_CXXCXGKG; 1.  
KW Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;  
FT Complete proteome.  
FT DOMAIN 5 69 J-DOMAIN.  
FT DOMAIN 76 117 GLY-RICH.  
FT REPEAT 146 153 CXXCXGKG MOTIF.  
FT REPEAT 163 170 CXXCXGKG MOTIF.  
FT REPEAT 189 196 CXXCXGKG MOTIF.  
FT REPEAT 203 210 CXXCXGKG MOTIF.  
FT METAL 146 146 ZINC 1 (BY SIMILARITY).  
FT METAL 149 149 ZINC 1 (BY SIMILARITY).  
FT METAL 163 163 ZINC 2 (BY SIMILARITY).  
FT METAL 166 166 ZINC 2 (BY SIMILARITY).  
FT METAL 189 189 ZINC 2 (BY SIMILARITY).  
FT METAL 192 192 ZINC 2 (BY SIMILARITY).  
FT METAL 203 203 ZINC 1 (BY SIMILARITY).  
FT METAL 206 206 ZINC 1 (BY SIMILARITY).  
SQ SEQUENCE 376 AA; 41020 MW; CA8631D6FA120C24 CRC64;  
  
Query Match 1.9%; Score 7; DB 1; Length 376;  
Best Local Similarity 100.0%; Pred.No. 48;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 235 GGGAGGG 241  
Db 78 GGGAGGG 84  
|||||  
78 GGGAGGG 84  
  
RESULT 44  
DNAJ LISMO  
ID DNAJ LISMO STANDARD; PRT; 377 AA.  
AC Q955A3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Chaperone protein dnaJ.  
GN DNAJ OR LM01472.  
OS Listeria monocytogenes  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=10403S;  
RC MEDLINE=20163771; PubMed=10701836;  
RA Hanawa T., Kai M., Kamiya S., Yamamoto T.;  
RT "Cloning, sequencing, and transcriptional analysis of the dnaK heat  
RL shock operon of Listeria monocytogenes.";  
RN Cell Stress Chaperones 5:21-29(2000).  
RP SEQUENCE FROM N.A.  
RC STRAIN=EGD-e / Serovar 1/2a;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,



RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chatouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dusauguet O.,  
 RA Entian K.-D., Faihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Matournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordieck G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,  
 RT "Comparative genomics of *Listeria species*,"  
 RL Science 294:849-852(2001).  
 CC -1- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,  
 CC THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).  
 CC -1- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE DNAJ FAMILY.  
 CC -1- SIMILARITY: Contains 1 J domain.  
 CC -1- SIMILARITY: Contains 1 CR domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB023064; BAA82790.1; -;  
 DR EMBL; AL591979; CAC99550.1; -;  
 DR PIR; AH1258; AH1258.  
 DR PIR; T43739; T43739.  
 DR HSSP; P08622; IBOZ.  
 DR ListIlist; LMO01472; -;  
 DR InterPro; IPR02939; DnaJ\_C.  
 DR InterPro; IPR001305; DnaJ\_CXXCXGKG.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR InterPro; IPR003095; Hsp\_DnaJ.  
 DR Pfam; PF00226; DnaJ; 1.  
 DR Pfam; PF01556; DnaJ\_C; 1.  
 DR Pfam; PF06884; DnaJ\_CXXCXGKG; 1.  
 DR PRINTS; PR00625; DNAJPROTEIN.  
 DR SMART; SM00271; DnaJ; 1.  
 DR PROSITE; PS0076; DnaJ\_1; 1.  
 DR PROSITE; PS0076; DnaJ\_2; 1.  
 DR PROSITE; PS00637; DnaJ\_CXXCXGKG; 1.  
 KW Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;  
 KW Complete proteome.  
 FT DOMAIN 5 69 J-DOMAIN.  
 FT REPEAT 76 117 GLY-RICH.  
 FT REPEAT 147 154 CXXCXGKG MOTIF.  
 FT REPEAT 164 171 CXXCXGKG MOTIF.  
 FT REPEAT 190 197 CXXCXGKG MOTIF.  
 FT REPEAT 204 211 CXXCXGKG MOTIF.  
 FT METAL 147 147 ZINC 1 (BY SIMILARITY).  
 FT METAL 150 150 ZINC 1 (BY SIMILARITY).  
 FT METAL 164 164 ZINC 2 (BY SIMILARITY).  
 FT METAL 167 167 ZINC 2 (BY SIMILARITY).  
 FT METAL 190 190 ZINC 2 (BY SIMILARITY).  
 FT METAL 193 193 ZINC 2 (BY SIMILARITY).  
 FT METAL 204 204 ZINC 1 (BY SIMILARITY).  
 FT METAL 207 207 ZINC 1 (BY SIMILARITY).  
 SQ SEQUENCE 377 AA; 41077 MW; 70A8C5836239310A CRC64;  
 Query Match 1.9%; Score 7; DB 1; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGG 241  
 Db 79 GGGAGG 85

RESULT 45  
 NLPD SALDU STANDARD; PRT; 377 AA.  
 ID NLPD SALDU STANDARD; PRT; 377 AA.  
 AC P39760;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Lipoprotein nlpD precursor.  
 GN NLPD.  
 OS Salmonella dublin.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=98360;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND INDUCTION.  
 RC STRAIN-Lane;  
 RX MEDLINE=99138748; PubMed=9973354;  
 RA Paesold G., Krause M.;  
 RT "Analysis of rpoS mRNA in Salmonella dublin: identification of  
 RT multiple transcripts with growth-phase dependent variation in  
 RT transcript stability,";  
 RL J. Bacteriol. 181:1264-1268(1999).  
 RN [2]  
 RP SEQUENCE OF 284-377 FROM N.A.  
 RC STRAIN-Lane;  
 RA Krause M.W., El-Gedaily A.;  
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY BE INVOLVED IN STATIONARY-PHASE SURVIVAL.  
 CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid  
 CC anchor (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ006131; CAA06881.1; -;  
 DR EMBL; X82129; CAA57639.1; -;  
 DR MEROPS; M37.UFW; -;  
 DR InterPro; IPR002482; LysM.  
 DR InterPro; IPR002886; Peptidase\_M37.  
 DR Pfam; PF01476; LysM; 1.  
 DR Pfam; PF01551; Peptidase\_M37; 1.  
 DR SMART; SM00257; LysM; 1.  
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.  
 KW Inner membrane; Lipoprotein; Repeat; Signal.  
 FT SIGNAL 1 25  
 FT CHAIN 26 377 LIPOPROTEIN NLPD.  
 FT LIPID 26 26 N-ACYL DIGLYCERIDE (BY SIMILARITY).  
 FT DOMAIN 203 250 4 X 7 AA APPROXIMATE REPEATS.  
 FT REPEAT 203 209 1.  
 FT REPEAT 225 231 2.  
 FT REPEAT 237 243 3.  
 FT REPEAT 244 250 4.  
 SQ SEQUENCE 377 AA; 39671 MW; DC7F59B943F86512 CRC64;  
 Query Match 1.9%; Score 7; DB 1; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 GSTGTSS 87  
 Db 348 GSTGTSS 354

RESULT 46  
 NLPD SALTY STANDARD; PRT; 377 AA.  
 ID NLPD SALTY STANDARD; PRT; 377 AA.  
 AC P40827;





CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; Z14000; CAA78389.1; -.  
 DR EMBL; AL031228; CAA20235.1; -.  
 DR PIR; A47380; A47380.  
 DR HSSP; P28990; 1CHC.  
 DR Genew; HGNC:10018; RING1.  
 DR MIM; 602045; -.  
 DR GO; GO:0005634; C:nucleus; IDA.  
 DR GO; GO:0016564; F:transcriptional repressor activity; IDA.  
 DR GO; GO:0008270; F:zinc ion binding activity; NAS.  
 DR GO; GO:0016568; P:chromatin modification; NAS.  
 DR InterPro; IPR001841; Znf\_Ring.  
 DR Pfam; PF00097; zf-C3HC4; 1.  
 DR SMART; SM00184; RING; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; 1.  
 DR PROSITE; PS00899; ZF\_RING\_2; 1.  
 KW Chromatin regulator; Nuclear protein; Transcription regulation;  
 KW Repressor; Zinc-finger.  
 FT ZN\_FING 19 59 RING-TYPE.  
 FT DOMAIN 176 231 GLY-RICH.  
 FT DOMAIN 285 348 GLY-RICH.  
 FT DOMAIN 172 175 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 SQ SEQUENCE 377 AA; 39145 MW; F406E93593E0CF69 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
 |||||  
 DB 307 GGGAGGG 313

## RESULT 48

DNJ2\_STRCO STANDARD; PRT; 378 AA.  
 ID DNJ2\_STRCO  
 AC Q9RDE7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Chapterone protein dnaJ2.  
 GN DUAJ2 OR SCO2554 OR SCC77.21C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Harris D.E., Quail M.A., Kleeer H.,  
 RA Thomson N.R., James K.D., Chaudhry D., Chen C.W., Collins M.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Hensby T., Howarth S.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Oliver K., O'Neil S.,  
 RA Huang C.-H., Kiese R., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,  
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 CC -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,  
 CC THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).  
 CC -!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.  
 CC -!- SIMILARITY: Contains 1 J domain.  
 CC -!- SIMILARITY: Contains 1 CR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AL939113; CAB66232.1; -.  
 DR HSSP; P08622; 1XBL.  
 DR InterPro; IPR002939; DnaJ\_C.  
 DR InterPro; IPR001305; DnaJ\_CXXCXGKG.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR InterPro; IPR003095; Hsp\_DnaJ.  
 DR Pfam; PF00226; DnaJ\_1.  
 DR Pfam; PF01556; DnaJ\_C; 1.  
 DR Pfam; PF00684; DnaJ\_CXXCXGKG; 1.  
 DR PRINTS; PR00625; DnaJ\_PROTEIN.  
 DR SMART; SM00271; DnaJ; 1.  
 DR PROSITE; PS00636; DnaJ\_1; 1.  
 DR PROSITE; PS0076; DnaJ\_2; 1.  
 KW Chapterone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;  
 KW Complete proteome.  
 FT DOMAIN 4 68 J-DOMAIN.  
 FT DOMAIN 75 113 GLY-RICH.  
 FT REPEAT 143 150 CXXCXGKG MOTIF.  
 FT REPEAT 160 167 CXXCXGKG MOTIF.  
 FT REPEAT 186 193 CXXCXGKG MOTIF.  
 FT REPEAT 200 207 CXXCXGKG MOTIF.  
 FT METAL 143 143 ZINC 1 (BY SIMILARITY).  
 FT METAL 146 146 ZINC 1 (BY SIMILARITY).  
 FT METAL 160 160 ZINC 2 (BY SIMILARITY).  
 FT METAL 163 163 ZINC 2 (BY SIMILARITY).  
 FT METAL 186 186 ZINC 2 (BY SIMILARITY).  
 FT METAL 189 189 ZINC 2 (BY SIMILARITY).  
 FT METAL 200 200 ZINC 1 (BY SIMILARITY).  
 FT METAL 203 203 ZINC 1 (BY SIMILARITY).  
 SQ SEQUENCE 378 AA; 40488 MW; 8C3DEF5B4CC6808D CRC64;

Query Match 1.9%; Score 7; DB 1; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGG 240  
 |||||  
 DB 75 AGGGAGG 81

## RESULT 49

NLPD\_ECOLI STANDARD; PRT; 379 AA.  
 ID NLPD\_ECOLI  
 AC P33648;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Lipoprotein nlpD precursor.  
 GN NLPD OR B2742 OR SF2765.  
 OS Escherichia coli and  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562, 623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=MP180;  
 RX MEDLINE=94179096; PubMed=8132457;  
 RA Ichikawa J.K., Li C., Fu J.C., Clarke S.;  
 RT "A gene at 59 minutes on the Escherichia coli chromosome encodes a  
 RT lipoprotein with unusual amino acid repeat sequences";  
 RL J. Bacteriol. 176:1630-1638(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=K12 / MG1655;

RA MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RA "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RN SEQUENCE OF 99-379 FROM N.A.  
RN SPECIES=E.coli; STRAIN=K12 / DH1;  
RC MEDLINE=94268497; PubMed=8208244;  
RX Takayanagi Y., Tanaka K., Takahashi H.;  
RA "Structure of the 5' upstream region and the regulation of the rpos  
RT gene of *Escherichia coli*.";  
RL Mol. Gen. Genet. 243:525-531(1994).  
RN [4]  
RN SEQUENCE FROM N.A.  
RN SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
RC MEDLINE=2272406; PubMed=12384590;  
RX Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RA "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
RT through comparison with genomes of *Escherichia coli* K12 and O157.";  
RL Nucleic Acids Res. 30:4432-4441(2002).  
CC - FUNCTION: MAY BE INVOLVED IN STATIONARY-PHASE SURVIVAL.  
CC - SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid  
CC anchor (Potential).  
CC - SIMILARITY: BELONGS TO THE E-COLI NLPD / HAEMOPHILUS LPPB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL; L07869; AAA17875.1; -;  
DR EMBL; U29579; AAA69252.1; -;  
DR EMBL; AE000358; AAC75784.1; -;  
DR EMBL; D17549; BAA04487.1; -;  
DR EMBL; AE015290; AAN44254.1; ALT\_INIT.  
DR PIR; B55522; B55522.  
DR MEROPS; M37.UFW; -;  
DR SWISS-2DPAGE; P33648; COLI.  
DR EcoGene; EG12111; nlpd.  
DR InterPro; IPR002482; Peptidase\_M37.  
DR InterPro; IPR002886; Peptidase\_M37.  
DR Pfam; PF01476; LysM; 1.  
DR Pfam; PF01551; Peptidase\_M37; 1.  
DR SMART; SM00257; LysM; 1.  
DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.  
KW Inner membrane; Lipoprotein; Repeat; Signal; Complete proteome.  
FT SIGNAL 1 25  
FT CHAIN 26 379 LIPOPROTEIN NLPD.  
FT LIPID 26 26 N-ACYL DIGLYCERIDE.  
FT DOMAIN 66 97 4 X 8 AA TANDEN REPEATS OF Q-Q-P-Q-I-Q-  
FT P-V.  
FT REPEAT 66 73 1-1.  
FT REPEAT 74 81 1-2 (APPROXIMATE).  
FT REPEAT 82 89 1-3.  
FT REPEAT 90 97 1-4 (APPROXIMATE).  
FT DOMAIN 205 252 4 X 7 AA APPROXIMATE REPEATS.  
FT REPEAT 205 211 2-1.  
FT REPEAT 227 233 2-2.  
FT REPEAT 239 245 2-3.  
FT REPEAT 246 252 2-4.  
FT CONFLICT 139 139 G -> A (IN REF. 3).  
SQ SEQUENCE 379 AA; 40149 MW; A8B6A2B8456105FE CRC64;

Query Match 1.9%; Score 7; DB 1; Length 379;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 81 GSTGTSS 87  
DB 350 GSTGTSS 356  
RESULT 50  
TCF7\_HUMAN STANDARD; PRT; 383 AA.  
ID TCF7\_HUMAN  
AC P36402; Q9UKI4;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Transcription factor 7 (T-cell-specific transcription factor 1) (TCF-  
DE 1) (T-cell factor 1).  
GN TCF7 OR TCF1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RN SEQUENCE FROM N.A. (ISOFORM 4L).  
RX MEDLINE=91114695; PubMed=1989880;  
RA van de Wetering M., Oosterwegel M.A., Dooijes D., Clevers H.;  
RT "Identification and cloning of TCF-1, a T lymphocyte-specific  
RT transcription factor containing a sequence-specific HMG box.";  
RL EMBO J. 10:123-132(1991).  
RN [2]  
RN SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
RX MEDLINE=92235082; PubMed=1569101;  
RA van de Wetering M., Oosterwegel M.A., Holstege P., Dooyes D.,  
RA Suljkerbuijk R., Geurts van Kessel A., Clevers H.;  
RT "The human T cell transcription factor-1 gene. Structure,  
RT localization, and promoter characterization.";  
RL J. Biol. Chem. 267:8530-8536(1992).  
RN [3]  
RN SEQUENCE OF 1-68 FROM N.A. (ISOFORM 4L), TISSUE SPECIFICITY, AND  
RP INDUCTION BY TCF7L2 AND CTNNB1.  
RX MEDLINE=99420370; PubMed=10489374;  
RA Roose J., Huls G., van Beest M., Moerer P., van der Horn K.,  
RA Goldschmeding R., Logtenberg T., Clevers H.;  
RT "Synergy between tumor suppressor APC and the beta-catenin-Tcf4 target  
RT Tcf1.";  
RL Science 285:1923-1926(1999).  
RN [4]  
RN UTILISATION OF AN UPSTREAM PROMOTER, AND IDENTIFICATION OF ISOFORMS  
IS AND 4L.  
RX MEDLINE=96182076; PubMed=8622675;  
RA van de Wetering M., Castrop J., Korinek V., Clevers H.;  
RT "Extensive alternative splicing and dual promoter usage generate Tcf-1  
RT protein isoforms with differential transcription control properties.";  
RL Mol. Cell. Biol. 16:745-752(1996).  
RN [5]  
RN INTERACTION WITH CTNNB1.  
RX MEDLINE=98147763; PubMed=9488439;  
RA Korinek V., Barker N., Willert K., Molenaar M., Roose J., Wagenaar G.,  
RA Markman M., Lammers W., Destree O., Clevers H.;  
RT "Two members of the Tcf family implicated in Wnt/b-catenin signaling  
RT during embryogenesis in the mouse.";  
RL Mol. Cell. Biol. 18:1248-1256(1998).  
RN [6]  
RN INTERACTION WITH CTNNB1 AND AES.  
RX MEDLINE=98454908; PubMed=9783587;  
RA Roose J., Molenaar M., Peterson J., Hurenkamp J., Brantjes H.,  
RA Moerer P., van de Wetering M., Destree O., Clevers H.;  
RT "The Xenopus Wnt effector XTCF-3 interacts with Groucho-related  
RT transcriptional repressors.";  
RL Nature 395:608-612(1998).  
RN [7]  
RN INTERACTION WITH AES; TLE1; TLE2; TLE3 AND TLE4.  
RP

RX MEDLINE=21169341; PubMed=11266540;  
 RA Brantjes H., Roese J., van De Wetering M., Clevers H.;  
 RT "All Tcf HMG box transcription factors interact with Groucho-related  
 RL co-repressors.";  
 CC Nucleic Acids Res. 29:1410-1419(2001).  
 CC -!- FUNCTION: Transcriptional activator involved in T-cell lymphocyte  
 CC differentiation. Necessary for the survival of CD4(+) CD8(+)   
 CC immature thymocytes. Isoforms lacking the N-terminal CTNNB1  
 CC binding domain cannot fulfill this role. Binds to the T-  
 CC lymphocyte-specific enhancer element (5-WCAGAAG-3') found in the  
 CC promoter of the CD3E gene. May also act as feedback  
 CC transcriptional repressor of CTNNB1 and TCF7L2 target genes (By  
 CC similarity). TLE1, TLE2, TLE3 and TLE4 repress transactivation  
 CC mediated by TCF7 and CTNNB1.  
 CC -!- SUBUNIT: Binds the armadillo repeat of CTNNB1 and forms a stable  
 CC complex. Interacts with AEB, TLE1, TLE2, TLE3 and TLE4.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative promoter;  
 CC Comment=2 series of isoforms, L (shown here) and S, are  
 CC produced by use of alternative promoters;  
 CC Event=Alternative splicing; Named isoforms=8;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=4L; Synonyms=B;  
 CC IsoId=P36402-1; Sequence=Displayed;  
 CC Name=4S;  
 CC IsoId=P36402-2; Sequence=VSP\_006961;  
 CC Name=1L;  
 CC IsoId=P36402-3; Sequence=VSP\_006960;  
 CC Note=Produced by alternative splicing of isoform 4L;  
 CC Name=1S; Synonyms=B;  
 CC IsoId=P36402-4; Sequence=VSP\_006961; VSP\_006960;  
 CC Note=Produced by alternative splicing of isoform 4S;  
 CC Name=2L;  
 CC IsoId=P36402-5; Sequence=VSP\_002191;  
 CC Note=Produced by alternative splicing of isoform 4L;  
 CC Name=2S; Synonyms=B;  
 CC IsoId=P36402-6; Sequence=VSP\_006961; VSP\_002191;  
 CC Note=Produced by alternative splicing of isoform 4S;  
 CC Name=3L;  
 CC IsoId=P36402-7; Sequence=VSP\_002192;  
 CC Note=Produced by alternative splicing of isoform 4L;  
 CC Name=3S; Synonyms=C;  
 CC IsoId=P36402-8; Sequence=VSP\_006961; VSP\_002192;  
 CC Note=Produced by alternative splicing of isoform 4S;  
 CC -!- TISSUE SPECIFICITY: Predominantly in T-cells. Also detected in  
 CC proliferating intestinal epithelial cells and in the basal  
 CC epithelial cells of mammary gland epithelium.  
 CC -!- INDUCTION: By TCF7L2 and CTNNB1.  
 CC -!- SIMILARITY: Belongs to the TCF/Lef family.  
 CC -!- SIMILARITY: Contains 1 HMG box domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59869; CAA42526.1; -;  
 DR EMBL; X59870; CAA42527.1; -;  
 DR EMBL; X59871; CAA42528.1; -;  
 DR EMBL; X63901; CAA56795.1; -;  
 DR EMBL; Z47362; CAA87440.1; -;  
 DR EMBL; AF163776; AAF00616.1; -;  
 DR PIR; A38095; A38095.  
 DR PIR; B38095; B38095.  
 DR PIR; C38095; C38095.  
 DR HGSP; P27782; 2LEF.  
 DR TRANSFAC; T01001; -;  
 DR Genew; HGNC:11639; TCF7.

DR MTM; 189908; -;  
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.  
 DR GO; GO:0006955; P:immune response; TAS  
 DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.  
 DR InterPro; IPR000910; HMG\_box; 1.  
 DR Pfam; PF00505; HMG\_box; 1.  
 DR SMART; SM00398; HMG; 1.  
 DR PROSITE; PS01118; HMG\_BOX\_2; 1.  
 KW Transcription regulation; Activator; Repressor; Trans-acting factor;  
 KW Nuclear protein; DNA-binding; Wnt signaling pathway;  
 KW Alternative splicing; Alternative promoter usage.  
 FT DOMAIN 1 58 CTNNB1 BINDING.  
 FT DNA BIND 268 336 HMG\_BOX.  
 FT DOMAIN 343 347 Missing (in isoform 2S, isoform 3S and  
 FT VARSPLIC 1 114 isoform 4S).  
 FT FTID=VSP\_006961.  
 FT ETNPREDKDGQESLSMSSSSSPA -> DPGSPKKCRAR  
 FT FGLNQOTDMGQCRKKKCIYLPGEGRCPSPVSDSALG  
 FT CPSPAPQSPSVHLLPRPPTELLTSPAERHLHPQVSPLLS  
 FT ASQPGPHRPPAPCRHAYSNRLDRWPSRHTPGRLQE  
 FT PTP (in isoform 1L and isoform 1S).  
 FT /FTID=VSP\_006960.  
 FT ETNPREDKDGQESLSMSSSSSPA -> GGRNARGTYP  
 FT EKAAPAPFLPMTVL (in isoform 2L and  
 FT isoform 2S).  
 FT /FTID=VSP\_002191.  
 FT ETNPREDKDGQESLSMSSSSSPA -> DPGSPKKCRAR  
 FT FGLNQOTDMGQPCR (in isoform 3L and isoform  
 FT 3S).  
 FT /FTID=VSP\_002192.  
 FT D -> N (IN REF. 3).  
 FT CONFLICT 21 21 A -> AA (IN REF. 3).  
 FT CONFLICT 41 41 S -> G (IN REF. 3).  
 FT CONFLICT 64 64  
 FT S SEQUENCE 383 AA; 41685 MW; 12C1972B7614E18B CRC64;  
 SQ  
 Query Match 1.9%; Score 7; DB 1; Length 383;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 235 GGGAGGG 241  
 Db 8 GGGAGGG 14  
 RESULT 51  
 ID RO32\_XENLA STANDARD; PRT; 385 AA.  
 AC P51932;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Heterogeneous nuclear ribonucleoprotein A3 homolog 2 (hnRNP A3(B)).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93197168; PubMed=8451200;  
 RA Good P.J., Rebert M.L., David I.B.;  
 RT "Three new members of the RNP protein family in Xenopus.";  
 RL Nucleic Acids Res. 21:999-1006(1993).  
 CC -!- SUBCELLULAR LOCATION: Nuclear; component of ribonucleosomes (By  
 CC similarity).  
 CC -!- SIMILARITY: BELONGS TO THE A/B GROUP OF HNRNP, WHICH ARE BASIC AND  
 CC GLY-RICH PROTEINS.  
 CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59869; CAA42526.1; -;  
 DR EMBL; X59870; CAA42527.1; -;  
 DR EMBL; X59871; CAA42528.1; -;  
 DR EMBL; X63901; CAA56795.1; -;  
 DR EMBL; Z47362; CAA87440.1; -;  
 DR EMBL; AF163776; AAF00616.1; -;  
 DR PIR; A38095; A38095.  
 DR PIR; B38095; B38095.  
 DR PIR; C38095; C38095.  
 DR HGSP; P27782; 2LEF.  
 DR TRANSFAC; T01001; -;  
 DR Genew; HGNC:11639; TCF7.

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L02957; AAA49950.1; --  
 DR PIR; S40778; S40778.  
 DR HSSP; P09651; LUPL.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 2.  
 DR SMART; SM00360; RRM; 2.  
 DR PROSITE; PS0102; RRM; 2.  
 DR PROSITE; PS00030; RRM RNP 1; 2.  
 KW Nuclear protein; RNA-Binding; Ribonucleoprotein; Repeat.  
 FT DOMAIN 27 110 RNA-BINDING (RRM) 1.  
 FT METAL 118 206 RNA-BINDING (RRM) 2.  
 FT DOMAIN 207 385 GLY-RICH.  
 SQ SEQUENCE 385 AA; 39438 MW; 96895F1B997C9E9D CRC64;

Query Match 1.9%; Score 7; DB 1; Length 385;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
 |||||  
 DB 302 GGGAGGG 308

RESULT 52  
 ID\_DNAJ\_HALCU STANDARD; PRT; 389 AA.  
 AC O34135;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chaperone protein dnaJ (Heat shock protein 40).  
 GN DNAS OR HSP40.  
 OS Halobacterium cutirubrum.  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OX NCBI\_TaxID=2242;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33170;  
 RX MEDLINE=971383250; PubMed=9236279;  
 RA Bustard K., Gupta R.S.;

RT "The sequences of heat shock protein 40 (DnaJ) homologs provide  
 evidence for a close evolutionary relationship between the  
 Deinococcus-thermus group and cyanobacteria.";  
 RL J. Mol. Evol. 45:193-205(1997).  
 CC -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,  
 THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).  
 CC -!- COPACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.  
 CC -!- SIMILARITY: Contains 1 J domain.  
 CC -!- SIMILARITY: Contains 1 CR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U93357; AAB96891.1; --  
 DR PIR; T44957; T44957.  
 DR HSSP; P08622; 1BQZ.  
 DR InterPro; IPR002939; DnaJ C.  
 DR InterPro; IPR001305; DnaJ\_CXXCXGKG.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR InterPro; IPR003095; Hsp\_DnaJ.

DR Pfam; PF00226; DnaJ; 1.  
 DR Pfam; PF01556; DnaJ C; 1.  
 DR Pfam; PF00684; DnaJ\_CXXCXGKG; 1.  
 DR PRINTS; PR00625; DNAJPROTEIN.  
 DR SMART; SM00271; DnaJ; 1.  
 DR PROSITE; PS00636; DnaJ 1; FALSE\_NEG.  
 DR PROSITE; PS00076; DnaJ 2; 1.  
 DR PROSITE; PS00637; DnaJ\_CXXCXGKG; FALSE\_NEG.

KW Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.  
 FT DOMAIN 4 68 J-DOMAIN.  
 FT DOMAIN 79 131 GLY-RICH.  
 FT REPEAT 165 172 CXXCXGKG MOTIF.  
 FT REPEAT 182 189 CXXCXGKG MOTIF.  
 FT REPEAT 208 215 CXXCXGKG MOTIF.  
 FT REPEAT 222 229 CXXCXGKG MOTIF.  
 FT METAL 165 165 ZINC 1 (BY SIMILARITY).  
 FT METAL 168 168 ZINC 1 (BY SIMILARITY).  
 FT METAL 182 182 ZINC 2 (BY SIMILARITY).  
 FT METAL 185 185 ZINC 2 (BY SIMILARITY).  
 FT METAL 208 208 ZINC 2 (BY SIMILARITY).  
 FT METAL 211 211 ZINC 2 (BY SIMILARITY).  
 FT METAL 222 222 ZINC 1 (BY SIMILARITY).  
 FT METAL 225 225 ZINC 1 (BY SIMILARITY).  
 SQ SEQUENCE 389 AA; 41390 MW; 12C1426CBF309633 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
 |||||  
 DB 103 GGGAGGG 109

RESULT 53  
 ID\_DNAJ\_HALN1 STANDARD; PRT; 391 AA.  
 AC Q9HRV3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chaperone protein dnaJ.  
 GN DNAS OR VNG0489G.  
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OX NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., LaSkey S.R., Balliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Madlocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlshocher M., Spudis J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;

RT "Genome sequence of Halobacterium species NRC-1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,  
 THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).  
 CC -!- COPACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.  
 CC -!- SIMILARITY: Contains 1 J domain.  
 CC -!- SIMILARITY: Contains 1 CR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
CC ENBL; AE005002; AAG19025.1; --
CC PIR; E84207; E84207.
CC HSP; P08622; BQZ.
CC InterPro; IPR002939; DnaJ_C.
CC InterPro; IPR001305; DnaJ_CXXCXGKG.
CC InterPro; IPR001623; DnaJ_N.
CC InterPro; IPR003095; Hsp_DnaJ.
CC Pfam; PF002226; DnaJ_1.
CC Pfam; PF01556; DnaJ_C; 1.
CC Pfam; PF00684; DnaJ_CXXCXGKG; 1.
CC PRINTS; P00625; DnaJPROTEIN.
CC SMART; SM00271; DnaJ; 1.
CC PROSITE; PS00636; DnaJ_1; FALSE_NEG.
CC PROSITE; PS00076; DnaJ_2; 1.
CC PROSITE; PS00637; DnaJ_CXXCXGKG; FALSE_NEG.
KW Chapterone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;
KW Complete proteome.
FT DOMAIN 4 68 J-DOMAIN.
FT DOMAIN 79 131 GLY-RICH.
FT REPEAT 165 172 CXXCXGKG MOTIF.
FT REPEAT 182 189 CXXCXGKG MOTIF.
FT REPEAT 208 215 CXXCXGKG MOTIF.
FT REPEAT 222 229 CXXCXGKG MOTIF.
FT METAL 165 165 ZINC 1 (BY SIMILARITY).
FT METAL 168 168 ZINC 1 (BY SIMILARITY).
FT METAL 182 182 ZINC 2 (BY SIMILARITY).
FT METAL 185 185 ZINC 2 (BY SIMILARITY).
FT METAL 208 208 ZINC 2 (BY SIMILARITY).
FT METAL 211 211 ZINC 2 (BY SIMILARITY).
FT METAL 222 222 ZINC 1 (BY SIMILARITY).
FT METAL 225 225 ZINC 1 (BY SIMILARITY).
SQ SEQUENCE 391 AA; 41713 MW; 3FFD7E01B5EA25F CRC64;

Query Match 1.9%; Score 7; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241
DB 103 GGGAGGG 109
|||||
RESULT 54
PCL_ECTHA STANDARD; PRT; 391 AA.
AC P42516;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 4-coumarate-CoA ligase (EC 6.2.1.12) (4CL) (4-coumaroyl-CoA
DE synthase) (Fragment).
GN PCL.
OS Ectothiorhodospira halophila.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales;
OC Ectothiorhodospiraceae; Halorhodospira.
OX NCBI_TaxID=1053;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BN9626;
RX MEDLINE=96283619; PubMed=8670821;
RA Kort R., Hoff W.D., van West M., Kroon A.R., Hoffer S.M.,
RA Vlieg K.H., Crislaard W., van Beumen J.J., Heilingwerf K.J.;
RT "The xanthopsins: a new family of eubacterial blue-light
RT photoreceptors."
RL EMBO J. 15:3209-3218 (1996).
RN [2]
RP SEQUENCE OF 1-112 FROM N.A.
RC STRAIN=BN9626;
RX MEDLINE=95072006; PubMed=7981196;
RA Baca M., Borgstahl G.E., Boissinot M., Burke P.M., Williams D.R.,

```

```

RA Slater K.A., Getsoff E.D.;
RT "Complete chemical structure of photoactive yellow protein: novel
RT thioester-linked 4-hydroxycinnamyl chromophore and photocycle
RL Biochemistry 33:14369-14377 (1994).
CC -|- FUNCTION: CONVERTS P-COUMARIC ACID INTO P-COUMARYL COA. THIS IS
CC NECESSARY FOR THE ACTIVATION OF THE PHOTOACTIVE YELLOW PROTEIN
CC (PYP) CHROMOPHORE.
CC -|- CATALYTIC ACTIVITY: ATP + 4-coumarate + CoA = AMP + diphosphate +
CC 4-coumaroyl-CoA.
CC -|- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
CC ENBL; X98887; CAA67392.1; --
CC EMBL; U17017; AAG61736.1; --
CC PIR; C55993; C55993.
CC InterPro; IPR000873; AMP-bind.
CC Pfam; PF00501; AMP-binding; 2.
CC PROSITE; PS00455; AMP BINDING; FALSE_NEG.
KW Ligase; Phenylpropanoid metabolism.
PT NON TER 391 391
SQ SEQUENCE 391 AA; 42438 MW; 698321FB6EB2A07D CRC64;

Query Match 1.9%; Score 7; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 TSGSTGT 85
DB 104 TSGSTGT 110
|||||
RESULT 55
HXAA_HUMAN STANDARD; PRT; 393 AA.
AC P31260; O43370; O43605; Q15949;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Homeobox protein Hox-A10 (Hox-1H) (Hox-1.8) (PL).
GN HOXA10 OR HOX1H.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=91288229; PubMed=1676505;
RA Lowney P., Corral J.C., Detmer K., Lebeau M.M., Deaven L.,
RA Lawrence H.J., Largman C.;
RT "A human Hox 1 homeobox gene exhibits myeloid-specific expression of
RT alternative transcripts in human hematopoietic cells."
RL Nucleic Acids Res. 19:3443-3449 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Mi X., Winters J.L., Stevens D.B., Fleischman R.A.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Bradshaw H., Hinds K., Keppler D.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 301-383 FROM N.A.
RX MEDLINE=90046832; PubMed=2573064;
RA Shen W.-F., Largman C., Lowney P., Corral J.C., Detmer K.,

```

HAUSER C.A., SIMONITCH T.A., HACK F.M., LAWRENCE H.J.;  
"Lineage-restricted expression of homeobox-containing genes in human  
hematopoietic cell lines";  
Proc. Natl. Acad. Sci. U.S.A. 86:8536-8540 (1989).  
[5]  
SEQUENCE OF 319-384 FROM N.A.  
RX MEDLINE=90098876; PubMed=2574852;  
RA Acampora D., D'Esposito M., Faiella A., Pannese M., Migliaccio E.,  
RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;  
"The human HOX gene family";  
Nucleic Acids Res. 17:10385-10402 (1989).  
[6]  
SEQUENCE OF 332-370 FROM N.A.  
RX MEDLINE=94145486; PubMed=7906121;  
RA Castronovo V., Kusaka M., Charlot A., Gielen J., Sobel M.;  
"Homeobox genes: potential candidates for the transcriptional control  
of the transformed and invasive phenotype";  
Biochem. Pharmacol. 47:137-143 (1994).  
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
CC BINDS TO THE DNA SEQUENCE 5'-AA[AT]TTTATTAC-3'.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1; Synonyms=Pl1;  
CC IsoId=P31260-1; Sequence=Displayed;  
CC Name=2; Synonyms=Pl2;  
CC IsoId=P31260-2; Sequence=VSP 002384, VSP 002385;  
CC -!- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.  
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A NUMBER OF  
CC FRAMESHIFTS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X58430; CAB86198.1; ALT FRAME.  
DR EMBL; AF040714; AAB96917.1; -.  
DR EMBL; AC004080; -; NOT ANNOTATED\_CDS.  
DR EMBL; M30599; AAA36006.1; -.  
DR EMBL; S69027; AAD14030.1; -.  
DR EMBL; S69029; AAD14031.1; -.  
DR PIR; I65265; I65265.  
DR PIR; S26402; S26402.  
DR PIR; S28814; S28814.  
DR HSSP; P02833; 9ANT.  
DR TRANSFAC; T01713; -.  
DR Genew; HGNC:5100; HOXA10.  
DR MIM; 142957; -.  
DR GO; GO:0007275; P:development; TAS.  
DR GO; GO:0007048; P:oncogenesis; TAS.  
DR GO; GO:0007283; P:spermatogenesis; TAS.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00024; HOMEBOX.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation; Alternative splicing.  
FT DOMAIN 26 39 POLY-GLY.  
FT DOMAIN 78 89 POLY-GLY.  
FT DOMAIN 116 120 POLY-PRO.  
FT DOMAIN 124 133 POLY-PRO.  
FT DOMAIN 208 215 POLY-GLY.  
FT DOMAIN 256 259 POLY-PRO.  
FT DNA\_BIND 319 378 HOMEBOX.

FT VARSPLIC 1 299 Missing (in isoform 2).  
FT VARSPLIC 300 302 /FTId=VSP 002384.  
FT VARSPLIC 300 302 DSL -> MCQ (in isoform 2).  
FT CONFLICT 1 6 MSCSES -> \_MFCTRNVSKGLSAPPAKLSHNNVMLGE  
(IN REF. 3).  
FT CONFLICT 69 69 G -> R (IN REF. 1).  
FT CONFLICT 136 137 QA -> HR (IN REF. 1).  
FT CONFLICT 332 332 L -> S (IN REF. 4).  
FT CONFLICT 334 334 L -> F (IN REF. 4).  
FT CONFLICT 370 370 R -> P (IN REF. 4).  
SQ SEQUENCE 393 AA; 40536 MW; 20F89542582D6F25 CRC64;  
Query Match 1.9%; Score 7; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 235 GGGAGGG 241  
DB 29 GGGAGGG 35  
|||||  
RESULT 56  
HEMI\_AQUAE STANDARD; PRT; 406 AA.  
ID HEMI\_AQUAE  
AC O67314;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GLUTR).  
GN HEMA OR AQ1279.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Shead M.A., Keller M., Anjay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
aeolicus";  
RL Nature 392:353-358 (1998).  
CC -!- CATALYTIC ACTIVITY: Glutamyl-tRNA (Glu) + NADPH = glutamate-1-  
CC semialdehyde + NADP(+) + tRNA (Glu).  
CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.  
CC -!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000732; AAC07274.1; -.  
DR PIR; G70410; G70410.  
DR HSSP; Q42843; 1B29.  
DR HAMAP; MF 00087; -; 1.  
DR InterPro; IPR000343; GLUTR.  
DR Pfam; PF05201; GLUTR\_N1.  
DR Pfam; PF05200; GLUTR\_NAD\_bind; 1.  
DR TIGRFAMs; TIGR01035; hema; 1.  
DR PROSITE; PS00747; GLUTR; 1.  
KW Porphyrin biosynthesis; Oxidoreductase; NADP; Complete proteome.  
FT ACT\_SITE 51 51 NUCLEOPHILE (BY SIMILARITY).  
FT ACT\_SITE 97 97 BASE (BY SIMILARITY).  
SQ SEQUENCE 406 AA; 46648 MW; 77F3ECC488FC4394 CRC64;  
Query Match 1.9%; Score 7; DB 1; Length 406;  
Best Local Similarity 100.0%; Pred. No. 52;



```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 354 ELAANYL 360
   |||||
Db 193 ELAANYL 199

RESULT 57
YB44_MYCPN
ID YB44_MYCPN STANDARD; PRT; 413 AA.
AC P75142;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Very hypothetical adhesin P1-like protein MPN144 (E07_orf413):
GN MPN144 OR MP010.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreid R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
  Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
  pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: BELONGS TO THE ADHESIN P1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000002; AAB95658.1; -.
DR PIR; S73336; S73336.
DR InterPro; IPR004940; Adhesin_P1.
DR Pfam; PF03257; Adhesin_P1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 413 AA; 43249 MW; 594375273CD5B6E5 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 413;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 235 GGGAGGG 241
   |||||
Db 387 GGGAGGG 393

RESULT 58
SO DROME
ID SO DROME STANDARD; PRT; 416 AA.
AC Q27350; Q9V4L0;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sine oculis protein.
GN SO OR C011121.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Head;
RX MEDLINE=95203683; PubMed=7896096;
RA Serikaku M.A., O'Tousa J.E.;

```

```

RT "Sine oculis is a homeobox gene required for Drosophila visual system
  development."
RL Genetics 138:1137-1150(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo, and Eye imaginal disk;
RX MEDLINE=94242440; PubMed=7910468;
RA Cheyette B.N.R., Green P.J., Martin K., Garren H., Hartenstein V.,
  Zipursky S.L.;
RT "The Drosophila sine oculis locus encodes a homeodomain-containing
  protein required for the development of the entire visual system."
RL Neuron 12:977-996(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
  Abril J.F., Agbayani A., An H.-J., Andrews-Frankkoch C., Miklos G.L.G.,
  Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Baldwin D.,
  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolehakov S.,
  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
  Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
  Chery J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
  de Fabros B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
  Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
  Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
  Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
  Hosoin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
  Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
  Shue B.C., Siden-Kiamos I., Simpsons M., Skupski M.P., Smith T.,
  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
  Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
  Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: REQUIRED FOR VISUAL SYSTEM DEVELOPMENT. MAY
  TRANSCRIPTIONALLY REGULATE GENES NECESSARY FOR OPTIC LOBE
  INVAGINATION AND BOLWIG'S NERVE FORMATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: IN DEVELOPING EMBRYOS, EXPRESSED IN THE EYE
  DISK EPITHELIUM, BOLWIG'S ORGAN AND THE OPTIC LOBE PRIMORDIUM AT
  AREAS OF INVAGINATION. IN ADULTS, PRESENT IN PHOTORECEPTOR CELLS
  IN THE APICAL REGIONS OF THE RETINA, AND IN OPTIC LOBES
CC -1- DEVELOPMENTAL STAGE: IN THE EYE IMAGINAL DISK, FIRST EXPRESSED AT
  THE ONSET OF THE THIRD INSTAR AND CONTINUES THROUGHOUT THIS STAGE.
  EXPRESSION IN THE OPTIC LOBE PRIMORDIUM BEGINS AT STAGE 5 AND
  DISAPPEARS WHEN INVAGINATION IS COMPLETED (STAGE 12). FURTHER
  EXPRESSION IS NOTED IN OPTIC LOBE GLANGIA IN LATE THIRD INSTAR.
CC -1- MISCELLANEOUS: SO MUTANTS SPECIFICALLY DISPLAY VISUAL DEFECTS.
  THESE RANGE FROM REDUCED OCELLI AND OMATIDIA NUMBER IN WEAK LOSS-
  OF-FUNCTION PHENOTYPES TO COMPLETE ABSENCE OF COMPOUND EYES AND
  BOLWIG'S ORGAN IN MORE SEVERE LETHAL PHENOTYPES.
CC -1- SIMILARITY: BELONGS TO THE SIX/SINE OCULIS HOMEOBOX FAMILY.

```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC EMBL; S77459; AAB34685.1; --  
 CC EMBL; L31626; AAA21800.1; --  
 CC EMBL; AE003841; AAF59260.1; --  
 CC PIR; S50342; S50342.  
 CC HSSP; P41778; 1DU6.  
 CC TRNGFAC; T03273; --  
 CC FlyBase: FBgn0003460; so.  
 CC GO; GO:0007823; P:circadian rhythm; IGI.  
 CC GO; GO:0008347; P:glia cell migration; IMP.  
 CC InterPro; IPR001356; Homeobox.  
 CC InterPro; IPR000047; HTH lambdarepressor.  
 CC InterPro; IPR007105; SIX.  
 CC InterPro; IPR007106; SIX\_SINE\_homeo.  
 CC Pfam; PF00046; homeobox; 1.  
 CC PRINTS; PR00031; HTHREPRESSR.  
 CC ProDom; PD000010; Homeobox; 1.  
 CC SMART; SM00389; HOX; 1.  
 CC PROSITE; PS00027; HOMEBOX 1; 1.  
 CC PROSITE; PS50071; HOMEBOX\_2; 1.  
 CC Developmental protein; Homeobox; DNA-binding; Nuclear protein.  
 KW DOMAIN 42 47 POLY-ASN.  
 FT DOMAIN 52 55 POLY-ASN.  
 FT DOMAIN 80 85 POLY-GLY.  
 FT DOMAIN 307 316 POLY-GLN (OPA REPEAT).  
 FT DOMAIN 366 372 POLY-GLY.  
 FT DOMAIN 375 379 POLY-ALA.  
 FT DNA\_BIND 218 277 HOMEBOX.  
 SQ SEQUENCE 416 AA; 45025 MW; 669E4BD95B857996 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 416;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241

Db 65 GGGAGGG 71

RESULT 59

PCL\_RHOCA

ID\_PCL\_RHOCA STANDARD; PRT; 417 AA.

AC O69140;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 4-coumarate--CoA ligase (EC 6.2.1.12) (4CL) (4-coumaroyl-CoA

DE synthase).

GN PCL.

OS Rhodospirillum rubrum (Rhodospirillum rubrum).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;

OC Rhodospirillaceae; Rhodospirillum.

OX NCBI\_TaxID=1061;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SB1003 / St. Louis;

RA Jiang Z., Bauer E.C.;

RT "Genetic characterization of photoactive yellow protein from

RT Rhodospirillum rubrum";

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: CONVERTS P-COUMARIC ACID INTO P-COUMARYL COA. THIS IS

CC NECESSARY FOR THE ACTIVATION OF THE PHOTOACTIVE YELLOW PROTEIN

CC (PYP) CHROMOPHORE (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: ATP + 4-coumarate + CoA = AMP + diphosphate +

CC 4-coumaroyl-CoA.

CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME

CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC EMBL; AF064095; AAC17429.1; --  
 CC InterPro; IPR000873; AMP-bind.  
 CC Pfam; PF00501; AMP-binding; 1.  
 CC PROSITE; PS00455; AMP BINDING; 1.  
 CC Ligase; phenylpropanoid metabolism.  
 KW SEQUENCE 417 AA; 44047 MW; B3F7BE333D705DBD1 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 417;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 TSGSTGT 85

Db 133 TSGSTGT 139

RESULT 60

CC38\_RAT

ID\_CC38\_RAT STANDARD; PRT; 421 AA.

AC Q8VRW5;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Voltage-dependent calcium channel gamma-8 subunit (Neuronal voltage-  
 gated calcium channel gamma-8 subunit).

GN CACNG8.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley;

RX MEDLINE=21601102; PubMed=11738816;

RA Chu P.-J., Robertson H.M., Best P.M.;

RT "Calcium channel gamma subunits provide insights into the evolution of

RT this gene family.";

RL Gene 280:37-48(2001).

CC -!- FUNCTION: THOUGHT TO STABILIZE THE CALCIUM CHANNEL IN AN

CC INACTIVATED (CLOSED) STATE (BY SIMILARITY).

CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FIVE SUBUNITS:

CC ALPHA-1, ALPHA-2/DELTA, BETA AND GAMMA.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).

CC -!- SIMILARITY: BELONGS TO THE PMP-22 / EMP / MP20 FAMILY. CACNG

CC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC EMBL; AF361346; AAL50041.1; --  
 CC InterPro; IPR004031; PMP22\_Claudin.  
 CC Pfam; PF00822; PMP22\_Claudin; 1.  
 CC Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
 KW Calcium channel.

FT TRANSMEM 19 39 POTENTIAL.

FT TRANSMEM 127 147 POTENTIAL.

FT TRANSMEM 157 177 POTENTIAL.

FT TRANSMEM 207 227 POTENTIAL.

```
FT TRANSMEM 318 338 POTENTIAL.
FT DOMAIN 245 360 GLY-RICH.
SQ SEQUENCE 421 AA; 43268 MW; 0FDB1461BDC780A9 CRC64;

Query Match
Best Local Similarity 1.9%; Score 7; DB 1; Length 421;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGG 240
DB 244 AGGAGG 250
|||||

RESULT 61
CC8_MOUSE STANDARD; PRT; 423 AA.
ID CCG8_MOUSE
AC Q8VHW2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Voltage-dependent calcium channel gamma-8 subunit (Neuronal voltage-
DE sated calcium channel gamma-8 subunit).
GN CACNG8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=21601102; PubMed=11738816;
RA Chu P.-J., Robertson H.M., Best P.M.;
RT "Calcium channel gamma subunits provide insights into the evolution of
RT this gene family.";
RL Gene 280:37-48(2001).
CC -!- FUNCTION: THOUGHT TO STABILIZE THE CALCIUM CHANNEL IN AN
CC INACTIVATED (CLOSED) STATE (BY SIMILARITY).
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FIVE SUBUNITS:
CC ALPHA-1, ALPHA-2/DELTA, BETA AND GAMMA.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PMP-22 / EMP / MP20 FAMILY. CACNG
CC SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; AF361350; AAL50045.1; -
DR MGD; MGI:1912376; CACNG8
DR InterPro; IPR004031; PMP22_Claudin.
DR Pfam; PF0822; PMP22_Claudin; 1.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel.
FT TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 207 227 POTENTIAL.
FT TRANSMEM 318 338 POTENTIAL.
FT TRANSMEM 245 360 GLY-RICH.
FT DOMAIN 245 360
SQ SEQUENCE 423 AA; 43453 MW; D3AD13E18F9EB02D CRC64;

Query Match
Best Local Similarity 1.9%; Score 7; DB 1; Length 423;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGG 240
DB 244 AGGAGG 250
|||||
```

```
RESULT 62
MKR2_SERQU STANDARD; PRT; 423 AA.
ID MKR2_SERQU
AC Q9DD48;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Makorin 2 (zinc finger protein YGHLC3HC4).
GN MKR2 OR YGHLC3HC4 OR YGHL2.
OS Seriola quinqueradiata (Five-ray yellowtail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Percoidae;
OC Carangidae; Seriola.
OX NCBI_TaxID=8161;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Takagi R., Saeki T., Abe S.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Gill, and Testis;
RX MEDLINE=21481893; PubMed=11597136;
RA Gray T.A., Azama K., Whitmore K., Min A., Abe S., Nicholls R.D.;
RT "Phylogenetic conservation of the makorin-2 gene, encoding a multiple
RT zine-finger protein, antisense to the raf1 proto-oncogene.";
RL Genomics 77:119-126(2001).
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: Contains 4 C3H1-type zinc fingers.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; AB047517; BAB18861.1; -
DR EMBL; AB049436; BAB18815.1; -
DR EMBL; AB049439; BAB39863.1; -
DR EMBL; AB049437; BAB39861.1; -
DR EMBL; AB049438; BAB39862.1; -
DR HSSP; P28950; 1CHC.
DR InterPro; IPR000571; Znf_CCCH.
DR InterPro; IPR001841; Znf_Ring.
DR Pfam; PF00097; zf-C3HC4; 1.
DR Pfam; PF00642; zf-CCCH; 4.
DR SMART; SM00184; RING; 1.
DR SMART; SM00356; Znf_C3H1; 4.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
KW Zinc-finger; Repeat.
FT ZN_FING 8 26 C3H1-TYPE 1.
FT ZN_FING 37 55 C3H1-TYPE 2.
FT ZN_FING 158 186 C3H1-TYPE 3.
FT DOMAIN 190 219 MAKORIN-TYPE CYS-HIS.
FT ZN_FING 235 289 RING-TYPE.
FT ZN_FING 324 344 C3H1-TYPE 4.
SQ SEQUENCE 423 AA; 46376 MW; APABF7AB14E86997 CRC64;

Query Match
Best Local Similarity 1.9%; Score 7; DB 1; Length 423;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGG 241
DB 74 GGGAGG 80
|||||

RESULT 63
```

CCG8 HUMAN  
 ID CCG8 HUMAN STANDARD; PRT; 425 AA.  
 AC Q8WX55; Q8BXT0; Q8BY23;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Voltage-dependent calcium channel gamma-8 subunit (Neuronal voltage-gated calcium channel gamma-8 subunit).  
 DE CACNG8 OR CACNG6.  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21601102; PubMed=11739816;  
 RA Chu P.-J., Robertson H.M., Best P.M.;  
 RT "Calcium channel gamma subunits provide insights into the evolution of this gene family.";  
 RT this gene family.";  
 RL Gene 280:37-48(2001).  
 RN [2]  
 RP SEQUENCE OF 12-426 FROM N.A.  
 RX MEDLINE=21100909; PubMed=11170751;  
 RA Burgess D.L., Geirides L.A., Foreman P.J., Noebels J.L.;  
 RT "A cluster of three novel Ca(2+) channel gamma subunit genes on chromosome 19q13.4: evolution and expression profile of the gamma subunit gene family.";  
 RL Genomics 71:339-350(2001).  
 RN [3]  
 RP SEQUENCE OF 1-204 FROM N.A.  
 RC TISSUE=Cerebellum;  
 RA Black J.L. III, Kryzer T.J., Lennon V.A.;  
 RT "Proposed Homo sapiens voltage-gated calcium channel gamma-6 subunit.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THOUGHT TO STABILIZE THE CALCIUM CHANNEL IN AN INACTIVATED (CLOSED) STATE (BY SIMILARITY).  
 CC -1- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FIVE SUBUNITS: ALPHA-1, ALPHA-2/DELTA, BETA AND GAMMA.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE PMP-22 / EMP / MP20 FAMILY. CACNG SUBFAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; AF361354; AAL50049.1; --  
 DR EMBL; AF288388; AAK20031.1; --  
 DR EMBL; AF234892; AAK15019.1; ALT\_INIT.  
 DR Genew; HGNC:13628; CACNG8.  
 DR MIM; 606900; --  
 DR GO; GO:0005891; C:voltage-gated calcium channel complex; NAS.  
 DR GO; GO:0005245; E:voltage-gated calcium channel activity; NAS.  
 DR GO; GO:0006816; P:calcium ion transport; NAS.  
 DR Pfan; PF00822; PMP22 Claudin; 1.  
 DR Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
 KW Calcium channel.  
 FT TRANSMEM 19 39  
 FT TRANSMEM 129 149  
 FT TRANSMEM 158 178  
 FT TRANSMEM 208 228  
 FT TRANSMEM 318 338  
 FT DOMAIN 247 380  
 FT GLY-RICH.  
 FT ME -> OV (IN REF. 3).  
 FT CONFLICT 1 2  
 FT CONFLICT 16 16  
 FT CONFLICT 67 70  
 FT CONFLICT 362 362  
 FT CONFLICT 386 386  
 FT CONFLICT 386 386

FT CONFLICT 394 395 RH -> AP (IN REF. 2).  
 FT CONFLICT 409 409 G -> E (IN REF. 2).  
 SQ SEQUENCE 425 AA; 43481 MW; E11105BDAF619D3D CRC64;  
 Query Match 1.9%; Score 7; DB 1; Length 425;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 234 AGGGAGG 240  
 DB 245 AGGGAGG 251  
 RESULT 64  
 SR54 THEAQ  
 ID SR54 THEAQ STANDARD; PRT; 429 AA.  
 AC 007347;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Signal recognition particle protein (Fifty-four homolog).  
 DE FFH.  
 GN Thermus aquaticus.  
 OS Bacteria; Deinococcus-Thermus; Deinococci; Thermaceae;  
 OC Thermus.  
 OX NCBI\_TaxID=271;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Freymann D.M., Keenan R.J., Stroud R.M., Walter P.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 1-294.  
 RX MEDLINE=97156020; PubMed=9002524;  
 RA Freymann D.M., Keenan R.J., Stroud R.M., Walter P.;  
 RT "Structure of the conserved GTPase domain of the signal recognition particle.";  
 RT Nature 385:361-364 (1997).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).  
 RX MEDLINE=98359118; PubMed=9695947;  
 RA Keenan R.J., Freymann D.M., Walter P., Stroud R.M.;  
 RT "Crystal structure of the signal sequence binding subunit of the signal recognition particle.";  
 RT Cell 94:181-191(1998).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF 1-294.  
 RX MEDLINE=99356018; PubMed=10426959;  
 RA Freymann D.M., Keenan R.J., Stroud R.M., Walter P.;  
 RT "Functional changes in the structure of the SRP GTPase on binding GDP and Mg2+GDP.";  
 RT Nat. Struct. Biol. 6:793-801(1999).  
 CC -1- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE RIBOSOMES.  
 CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A SMALL CYTOPLASMIC RNA (SC-RNA) MOLECULE AND PROTEIN FFH.  
 CC -1- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL SEQUENCE.  
 CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; U82109; AAB58502.1; --  
 DR PDB; 1FFH; 31-DEC-97.  
 DR PDB; 2FFH; 16-JUL-99.  
 DR PDB; 1NG1; 13-MAR-00.

DR PDB; 2NG1; 13-MAR-00.  
DR PDB; 3NG1; 13-MAR-00.  
DR PDB; 1JPU; 02-FEB-02.  
DR PDB; 1JPN; 02-FEB-02.  
DR PDB; 1LS1; 16-NOV-02.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR000897; SRP54.  
DR InterPro; IPR004125; SRP54\_SBP.  
DR InterPro; IPR004780; SRP\_sub.  
DR Pfam; PF00448; SRP54; 1.  
DR Pfam; PF02881; SRP54\_N; 1.  
DR Pfam; PF02978; SRP\_SBP; 1.  
DR ProDom; PD000819; SRP54; 1.  
DR SMART; SM00382; AAA; 1.  
DR TIGRFAMs; TIGR00959; fff; 1.  
DR PROSITE; PS00300; SRP54; 1.  
KW Signal recognition particle; GTP-binding; RNA-binding; 3D-structure.  
FT INIT MET 0  
FT DOMAIN 0  
FT DOMAIN 1 291 G-DOMAIN.  
FT NP\_BIND 292 429 M-DOMAIN.  
FT NP\_BIND 104 111 GTP (BY SIMILARITY).  
FT NP\_BIND 186 190 GTP (BY SIMILARITY).  
FT NP\_BIND 244 247 GTP (BY SIMILARITY).  
FT HELIX 2 13  
FT TURN 14 17  
FT HELIX 23 39  
FT TURN 40 41  
FT HELIX 44 59  
FT TURN 60 62  
FT HELIX 63 65  
FT HELIX 69 84  
FT TURN 85 86  
FT STRAND 98 103  
FT TURN 106 107  
FT HELIX 110 122  
FT TURN 123 125  
FT STRAND 128 132  
FT HELIX 138 151  
FT TURN 152 152  
FT STRAND 155 157  
FT TURN 160 161  
FT HELIX 164 177  
FT TURN 178 179  
FT STRAND 182 186  
FT HELIX 195 208  
FT STRAND 212 218  
FT HELIX 219 221  
FT TURN 222 223  
FT HELIX 224 235  
FT TURN 236 236  
FT STRAND 240 244  
FT HELIX 246 248  
FT HELIX 253 262  
FT STRAND 266 270  
FT HELIX 275 277  
FT STRAND 278 280  
FT HELIX 283 291  
FT TURN 292 292  
FT HELIX 299 305  
FT TURN 321 333  
FT TURN 334 334  
FT TURN 338 339  
FT HELIX 340 342  
FT TURN 343 343  
FT TURN 351 352  
FT HELIX 355 366  
FT TURN 367 367  
FT HELIX 370 374  
FT TURN 376 378  
FT HELIX 381 391  
FT TURN 392 392  
FT HELIX 395 414  
FT TURN 415 416

SQ SEQUENCE 429 AA; 47225 MW; 26D387A5A303EFBD CRC64;  
Query Match 1.9%; Score 7; DB 1; Length 429;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 326 QEAIGRL 332  
DB 9 QEAIGRL 15  
RESULT 65  
ELT2\_CABEL  
ID ELT2\_CABEL STANDARD; PRT; 433 AA.  
AC Q10655; Q18371;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Transcription factor elt-2.  
OS ELT-2 OR C33D3.1.  
GN Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=95301560; PubMed=7782329;  
RA Hawkins M.G., McGhee J.D.;  
RT "elt-2, a second GATA factor from the nematode Caenorhabditis  
elegans.";  
RL J. Biol. Chem. 270:14666-14671(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Sulston J.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBSJ databases.  
RN [3]  
RP REVISIONS.  
RA Sulston J.;  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBSJ databases.  
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF GENES CONTAINING THE GATA  
REGION.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- DEVELOPMENTAL STAGE: PRESENT AT ALL STAGES BUT HIGHEST IN  
EMBRYOS.  
CC -!- SIMILARITY: Contains 1 GATA-type zinc finger.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U25175; AAC316130.1; -  
CC EMBL; Z49867; CA990029.2; -  
CC FIR; A56953; A56953.  
CC HSP; P17678; IGAT.  
CC WormPep; C33D3.1; CE31430.  
CC InterPro; IPR000679; Znf\_GATA.  
CC Pfam; PF00320; GATA; 2.  
CC PRINTS; PR00619; GATAZNFINGER.  
CC SMART; SM00401; Znf\_GATA; 1.  
CC PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
CC PROSITE; PS0114; GATA\_ZN\_FINGER\_2; 1.  
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;  
Metal-binding; Nuclear protein.  
FT ZN\_FING 237 261 GATA-TYPE.  
SQ SEQUENCE 433 AA; 47116 MW; 59C2DEB075E5499 CRC64;  
Query Match 1.9%; Score 7; DB 1; Length 433;

Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 LYSIGIPV 189  
105 LYSIGIPV 111

Db

RESULT 66  
K1CM MOUSE  
ID K1CM MOUSE STANDARD; PRT; 437 AA.  
AC P08730;  
DT 01-AUG-1998 (Rel. 08, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Keratin, type I cytoskeletal 13 (Cytokeratin 13) (K13) (CK 13) (47 kDa  
cytokeratin).  
GN KRT13 OR KRT13-13.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV;  
RX MEDLINE=95213021; PubMed=7535287;  
RA Fillon M., Sarafian V., Luesier M., Belanger C., Lapointe L.,  
RA Royal A.;  
RT "Arrangement of a cluster of three mouse type I keratin genes  
expressed sequentially during esophageal-type epithelial cell  
differentiation";  
RL Genomics 24:303-310 (1994).  
RN [2]  
RP SEQUENCE OF 126-437 FROM N.A.  
RX MEDLINE=86120369; PubMed=2418416;  
RA Knapp B., Rentrop M., Schweizer J., Winter H.;  
RT "Nonepidermal members of the keratin multigene family: cDNA sequences  
and in situ localization of the mRNAs";  
RL Nucleic Acids Res. 14:751-763 (1986).  
RN [3]  
RP SEQUENCE OF 1-157 FROM N.A.  
RX MEDLINE=90323435; PubMed=1695590;  
RA Winter H., Rentrop M., Nischt R., Schweizer J.;  
RT "Tissue-specific expression of murine keratin K13 in internal  
stratified squamous epithelia and its aberrant expression during  
two-stage mouse skin carcinogenesis is associated with the  
methylation state of a distinct CpG site in the remote 5'-flanking  
region of the gene";  
RL Differentiation 43:105-114 (1990).  
CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
CC -1- MISCELLANEOUS: There are two types of cytoskeletal and  
microfibrillar keratin, I (acidic) and II (neutral to basic) (40-  
55 and 56-70 kDa, respectively).  
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U13921; AAC52150.1; -;  
DR EMBL; X03492; CAA27208.1; -;  
DR EMBL; X53320; CAA37407.1; -;  
DR PIR; A55682; A55682.  
DR MGD; MGI:101925; Krt11-13.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_I.  
DR Pfam; PF00038; filament\_1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PROSITE; PS00226; IF; 1.

Intermedate filament; Coiled coil; Keratin.  
HEAD.  
ROD.  
TAIL.  
LINKER 1.  
LINKER 12.  
COIL 1.  
COIL 2.  
COIL 2.  
SEQUENCE 437 AA; 47754 MW; 924128DBAC317E2F CRC64;

Query Match 1.9%; Score 7; DB 1; Length 437;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241  
54 GGGAGGG 60

Db

RESULT 67  
FXGA CHICK  
ID FXGA CHICK STANDARD; PRT; 440 AA.  
AC Q98937;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Forkhead box protein G1A (Forkhead-related protein FKHL2)  
DE (Transcription factor BF-2) (Brain factor 2) (BF2) (CBF-2) (T-14-6).  
GN FOXG1A OR FKHL2 OR HFHBF2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=White leghorn; TISSUE=Retina;  
RX MEDLINE=96338226; PubMed=8757134;  
RA Yuasa J., Hirano S., Yamagata M., Noda M.;  
RT "Visual projection map specified by topographic expression of  
transcription factors in the retina";  
RL Nature 382:632-635 (1996).  
CC -1- FUNCTION: MAY DETERMINE THE NASOTEMPORAL AXIS OF THE RETINA, AND  
CONSEQUENTLY SPECIFY THE TOPOGRAPHICAL PROJECTION OF THE RETINAL  
GANGLION-CELL AXONS TO THE TECTUM BY CONTROLLING EXPRESSION OF  
THEIR TARGET GENES.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -1- TISSUE SPECIFICITY: Retina.  
CC -1- DEVELOPMENTAL STAGE: CAN BE DETECTED IN REGIONS INCLUDING  
PRIMORDIAL RETINA AND NEUROEPITHELIUM BY EMBRYONIC DAY 2 (E2). AT  
E3, EXPRESSED IN THE TEMPORAL RETINA AND ASSOCIATED PIGMENT  
EPITHELIUM AS WELL AS IN PART OF THE DIENCEPHALON, AND AT E7 IS  
EXPRESSED IN RETINAL GANGLION CELLS. LEVELS BEGIN TO DECLINE FROM  
E4 AND ALMOST DISAPPEAR BY E10.  
CC -1- SIMILARITY: Contains 1 fork-head domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U47276; AAB08467.1; -;  
DR PIR; S71795; S71795.  
DR HSSP; Q63245; 2HFH.  
DR InterPro; IPR001766; TF Fork\_head.  
DR Pfam; PF00250; Fork head; 1.  
DR PRINTS; PR00053; FORKHEAD.  
DR ProDom; PD000425; TF Fork\_head; 1.  
DR SMART; SM00339; FH; 1.

Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 LYSIGIPV 189  
105 LYSIGIPV 111

Db

RESULT 66  
K1CM MOUSE  
ID K1CM MOUSE STANDARD; PRT; 437 AA.  
AC P08730;  
DT 01-AUG-1998 (Rel. 08, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Keratin, type I cytoskeletal 13 (Cytokeratin 13) (K13) (CK 13) (47 kDa  
cytokeratin).  
GN KRT13 OR KRT13-13.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV;  
RX MEDLINE=95213021; PubMed=7535287;  
RA Fillon M., Sarafian V., Luesier M., Belanger C., Lapointe L.,  
RA Royal A.;  
RT "Arrangement of a cluster of three mouse type I keratin genes  
expressed sequentially during esophageal-type epithelial cell  
differentiation";  
RL Genomics 24:303-310 (1994).  
RN [2]  
RP SEQUENCE OF 126-437 FROM N.A.  
RX MEDLINE=86120369; PubMed=2418416;  
RA Knapp B., Rentrop M., Schweizer J., Winter H.;  
RT "Nonepidermal members of the keratin multigene family: cDNA sequences  
and in situ localization of the mRNAs";  
RL Nucleic Acids Res. 14:751-763 (1986).  
RN [3]  
RP SEQUENCE OF 1-157 FROM N.A.  
RX MEDLINE=90323435; PubMed=1695590;  
RA Winter H., Rentrop M., Nischt R., Schweizer J.;  
RT "Tissue-specific expression of murine keratin K13 in internal  
stratified squamous epithelia and its aberrant expression during  
two-stage mouse skin carcinogenesis is associated with the  
methylation state of a distinct CpG site in the remote 5'-flanking  
region of the gene";  
RL Differentiation 43:105-114 (1990).  
CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
CC -1- MISCELLANEOUS: There are two types of cytoskeletal and  
microfibrillar keratin, I (acidic) and II (neutral to basic) (40-  
55 and 56-70 kDa, respectively).  
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U13921; AAC52150.1; -;  
DR EMBL; X03492; CAA27208.1; -;  
DR EMBL; X53320; CAA37407.1; -;  
DR PIR; A55682; A55682.  
DR MGD; MGI:101925; Krt11-13.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_I.  
DR Pfam; PF00038; filament\_1.  
DR PRINTS; PR01248; TYPE1KERATIN.  
DR PROSITE; PS00226; IF; 1.

```
DR PROSITE; PS00657; FORK HEAD_1; 1.
DR PROSITE; PS00658; FORK HEAD_2; 1.
DR PROSITE; PS00339; FORK HEAD_3; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
FT DEVELOPMENTAL PROTEIN.
FT DOMAIN 39 43 POLY-ARG.
FT DOMAIN 49 58 POLY-GLU.
FT DOMAIN 77 88 POLY-GLY.
FT DOMAIN 109 112 POLY-ALA.
FT DOMAIN 113 137 POLY-GLY.
FT DNA BIND 143 237 FORK-HEAD.
FT DOMAIN 321 330 POLY-PRO.
SQ SEQUENCE 440 AA; 44669 MW; 82266C3E1E103A48 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 440;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241
DB 86 GGGAGGG 92

RESULT 68
SOX4_MOUSE
ID SOX4_MOUSE STANDARD; PRT; 440 AA.
AC Q06831;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor SOX-4.
OS SOX4 OR SOX-4.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94008932; PubMed=8404853;
RA van de Wetering M., Oosterwegel M., van Norren K., Clevers H.C.;
RT "Sox-4, an Sry-like HMG box protein, is a transcriptional activator
in lymphocytes.";
RL EMBO J. 12:3847-3854 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93261838; PubMed=8493110;
RA Schilham M.W., van Eijk M., van de Wetering M., Clevers H.C.;
RT "The murine Sox-4 protein is encoded on a single exon.";
RL Nucleic Acids Res. 21:2009-2009 (1993).
RN [3]
RP SEQUENCE OF 69-122 FROM N.A.
RC STRAIN=Swiss Webster; TISSUE=Embryonic tooth;
RX MEDLINE=97079683; PubMed=8921394;
RA Stock D.W., Buchanan A.V., Zhao Z., Weiss K.M.;
RT "Numerous members of the Sox family of HMG box-containing genes are
expressed in developing mouse teeth.";
RL Genomics 37:234-237 (1996).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS WITH HIGH AFFINITY
TO THE T-CELL ENHANCER MOTIF 5'-AACAAAG-3' MOTIF.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOCYTES AND IN MOLAR AND
INCISOR TOOTH GERMS.
CC -1- SIMILARITY: Contains 1 HMG box domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
EMBL; X70298; CAA49779.1; -.

DR PROSITE; U70440; AAC52858.1; -.
DR PIR; S37303; S37303.
DR HSSP; P48436; 1SX9.
DR TRANSFAC; T01838; -.
DR MGD; MGI:98366; Sox4.
DR InterPro; IPR000910; HMG_12_box.
DR PIRam; PF00505; HMG_box_1.
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS01118; HMG_BOX_2; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; Activator.
FT DNA BIND 59 127 HMG_BOX.
FT DOMAIN 347 363 POLY-SER.
SQ SEQUENCE 440 AA; 45167 MW; 60BBE7FBF6FFBC21 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 440;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 240
DB 168 AGGAGGG 174

RESULT 69
SOX11_HUMAN
ID SOX11_HUMAN STANDARD; PRT; 441 AA.
AC P35716;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor SOX-11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96115609; PubMed=8666406;
RA Jay P., Goze C., Marsollier C., Taviaux S., Hardelin J.-P.,
Koopman P., Berta P.;
RT "The human SOX11 gene: cloning, chromosomal assignment and tissue
expression.";
RL Genomics 29:541-545 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20039622; PubMed=10574465;
RA Azuma T., Ao S., Saito Y., Yano K., Seki N., Wakao H., Masuho Y.,
Muramatsu M.;
RT "Human SOX11, an upregulated gene during the neural differentiation,
has a long 3' untranslated region.";
RL DNA Res. 6:357-360 (1999).
RN [3]
RP SEQUENCE OF 57-115 FROM N.A.
RX MEDLINE=93324385; PubMed=8332506;
RA Goze C., Poulat P., Berta P.;
RT "Partial cloning of SOX-11 and SOX-12, two new human SOX genes.";
RL Nucleic Acids Res. 21:2943-2943 (1993).
CC -1- FUNCTION: PROBABLY IMPORTANT IN THE DEVELOPING NERVOUS SYSTEM.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN THE NERVOUS SYSTEM.
CC -1- SIMILARITY: Contains 1 HMG box domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
EMBL; U23752; AAB08518.1; -.

```

```
DR EMBL; AB028641; BAA88122.1; -.
DR EMBL; X73038; CAA51519.1; -.
DR PIR; G01758; G01758.
DR PIR; S34118; S34118.
DR HSP; P48436; I8X9.
DR TRANSFAC; T02894; -.
DR Genew; HGNC:11191; SOX11.
DR MIM; 600898; -.
DR GO; GO:0007399; P:neurogenesis; TAS.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS0118; HMG_BOX 2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DNA BIND 49 117 HMG_BOX
FT DOMAIN 144 148 POLY-GLY.
FT DOMAIN 207 214 POLY-GLY.
FT DOMAIN 223 233 POLY-ASP.
FT DOMAIN 338 354 POLY-SER.
SQ SEQUENCE 441 AA; 46679 MW; 50567A654F503C63 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 441;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGG 240
Db 150 AGGGAGG 156
|||||

RESULT 70
COBB METJA
ID _COBB METJA STANDARD; PRT; 443 AA.
AC QS8816;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable cobyrinic acid A,C-diamide synthase.
GN COBB OR MJ1421
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=868087;
RA Buit C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake O., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii".
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: RESPONSIBLE FOR THE AMIDATION OF CARBOXYLIC GROUPS AT
CC POSITION A AND C OF EITHER COBYRINIC ACID OR HYDROGENOBIRINIC ACID.
CC NH(2) GROUPS ARE PROVIDED BY GLUTAMINE, AND ONE MOLECULE OF ATP
CC IS HYDROGENOLYZED FOR EACH AMIDATION (BY SIMILARITY).
CC -!- PATHWAY: Cobalamin biosynthesis.
CC -!- SIMILARITY: Belongs to the cobB/cobQ family. CobB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; U67583; AAB99432.1; -.
DR PIR; D64477; D64477.
DR TIGR; MJ1421; -.
DR HAMAP; MF_00027; -. 1.
DR InterPro; IPR004484; CbiA_P.
DR InterPro; IPR002586; CbiA_P.
DR Pfam; PF01656; CbiA; 1.
DR TIGRFAMs; TIGR00379; cobB; 1.
KW Cobalamin biosynthesis; Porphyrin biosynthesis; Complete proteome.
SQ SEQUENCE 443 AA; 50205 MW; EA866D0FC9ED212 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 443;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 IFNGKVL 53
Db 414 IFNGKVL 420
|||||

RESULT 71
GLNA HALVO
ID _GLNA HALVO STANDARD; PRT; 454 AA.
AC P43386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN GLNA.
OS Halobacterium volcanii (Haloflex volcanii).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Haloflex.
OX NCBI_TaxID=2246;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DS2;
RX MEDLINE=94365840; PubMed=7916055;
RA Brown J.R., Masuchi Y., Robb F.T., Doolittle W.F.;
RA "Evolutionary relationships of bacterial and archaeal glutamine
RA synthetase genes".
RT J. Mol. Evol. 38:566-576(1994).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

Query Match 1.9%; Score 7; DB 1; Length 454;  
Best Local Similarity 100.0%; Pred. No. 58;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 LESMGFD 338  
 |||||  
 Db 186 LESMGFD 192

RESULT 72  
 FXD1\_HUMAN STANDARD; PRT; 465 AA.  
 AC Q16676; Q12949;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Forkhead box protein D1 (Forkhead-related protein FKHL8) (Forkhead-  
 related transcription factor 4) (FREAC-4).  
 GN FOXD1 OR FKHL8 OR FREAC4.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96355467; PubMed=8702877;  
 RA Ennsson S., Pierrou S., Hulander M., Cederberg A., Hellqvist M.,  
 RA Carlsson P., Enerback S.;  
 RT "Characterization of the human forkhead gene FREAC-4. Evidence for  
 RT regulation by Wilms' tumor suppressor gene (WT-1) and p53.";  
 RL J. Biol. Chem. 271:21094-21099(1996).  
 RN [2]  
 RP SEQUENCE OF 120-225 FROM N.A.  
 RX MEDLINE=95045392; PubMed=7957066;  
 RA Pierrou S., Hellqvist M., Samuelsson L., Enerbaeck S., Carlsson P.;  
 RT "Cloning and characterization of seven human forkhead proteins:  
 RT binding site specificity and DNA bending.";  
 RL EMBO J. 13:5002-5012(1994).  
 CC -1- FUNCTION: BINDING OF FREAC-3 AND FREAC-4 TO THEIR COGNATE SITES  
 CC RESULTS IN BENDING OF THE DNA AT AN ANGLE OF 80-90 DEGREES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Contains 1 fork-head domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U59832; AAC50661.1; -;  
 CC EMBL; U59831; AAC50660.1; -;  
 CC EMBL; U13222; AAA92039.1; -;  
 CC PIR; G02738; G02738.  
 CC PIR; S51627; S51627.  
 CC HSP; Q63245; 2HFH.  
 CC TRANSFAC; T02472; -;  
 CC Genew; HGNC:3802; FOXD1.  
 CC MTM; 601091; -;  
 CC GO; GO:0003700; F:transcription factor activity; TAS.  
 CC InterPro; IPR001766; TF\_Fork\_head.  
 CC Pfam; PF00250; Fork\_head; 1.  
 CC PRINTS; PR00053; FORKHEAD.  
 CC ProDom; PD000425; TF\_Fork\_head; 1.  
 CC SMART; SM00339; FH; 1.  
 CC PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 CC PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 CC PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 CC Transcription regulation; DNA-binding; Nuclear protein.  
 FT DOMAIN 26 34 POLY-GLU.  
 FT DOMAIN 39 43 POLY-GLY.  
 FT DOMAIN 52 57 POLY-ARG.  
 FT DOMAIN 69 72 POLY-GLU.  
 FT DOMAIN 73 76 POLY-ASP.

FT DOMAIN 97 113 POLY-GLY.  
 FT DNA BIND 124 215 FORK-HEAD.  
 FT DOMAIN 231 234 POLY-ALA.  
 FT DOMAIN 252 256 POLY-ALA.  
 FT DOMAIN 259 266 POLY-PRO.  
 FT DOMAIN 293 303 POLY-ALA.  
 FT DOMAIN 309 315 POLY-PRO.  
 FT DOMAIN 375 390 POLY-ALA.  
 FT DOMAIN 428 434 POLY-ALA.  
 SQ SEQUENCE 465 AA; 46140 MW; D3E7854909CCBFAC CRC64;

Query Match 1.9%; Score 7; DB 1; Length 465;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GCGAGGG 241  
 |||||  
 Db 106 GCGAGGG 112

RESULT 73  
 FXD3\_MOUSE STANDARD; PRT; 465 AA.  
 AC Q61060;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Forkhead box protein D3 (HNF3/FH transcription factor gene) (HNF-2).  
 DE (Hepatocyte nuclear factor 3 forkhead homolog 2) (HFH-2).  
 GN FOXD3 OR HFH2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hromas R.A., Costa R.H., Xu D., Sutton J.L.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Labosky P.A., Kaetner K.H.;  
 RT "The winged helix transcription factor Hfh2 is expressed in neural  
 RT crest and spinal cord during mouse development.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Contains 1 fork-head domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U41047; AAA87569.1; -;  
 CC EMBL; AF067421; AAC28352.1; -;  
 CC HSP; Q63245; 2HFH.  
 CC TRANSFAC; T04166; -;  
 CC MGI; MGI:1347473; Foxd3.  
 CC InterPro; IPR001766; TF\_Fork\_head.  
 CC Pfam; PF00250; Fork\_head; 1.  
 CC PRINTS; PR00053; FORKHEAD.  
 CC ProDom; PD000425; TF\_Fork\_head; 1.  
 CC SMART; SM00339; FH; 1.  
 CC PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 CC PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 CC PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 CC DNA-binding; Nuclear protein; Transcription regulation.  
 FT DOMAIN 106 115 POLY-GLY.  
 FT DNA BIND 131 225 FORK-HEAD.  
 FT DOMAIN 252 257 POLY-ALA.



```

FT DOMAIN      265      270      POLY-ALA.
FT DOMAIN      275      281      POLY-ALA.
FT DOMAIN      380      399      POLY-GLY.
FT DOMAIN      447      457      POLY-ALA.
SQ SEQUENCE    465 AA; 47092 MW; 6F8B5B3D8C7564D CRC64;

Query Match      1.9%; Score 7; DB 1; Length 465;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241
    |||||
Db 385 GGGAGGG 391

RESULT 74
SYN-THETN      STANDARD;      PRT;      467 AA.
AC Q8RT3;
DT 28-FEB-2003 (Rel. 41, Last Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cysteine--trna synthetase (EC 6.1.1.16) (Cysteine--trna ligase)
DE (CYERS).
GN CYSS OR TTE2315.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
CC -!- CAVALYTIC ACTIVITY: ATP + L-cysteine + tRNA(Cys) = AMP +
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC Strong, to methionyl-tRNA synthetase.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE013173; AM25456.1; -.
DR HAMAP; MF_00041; -.
DR InterPro; IPR002308; Cys trna-synt 1a.
DR InterPro; IPR001412; trna-synt 1.
DR Pfam; PF01406; trna-synt 1e; 1.
DR TIGRFAMs; TIGR00435; cy85; 1.
DR PROSITE; PS00178; AA tRNA LIGASE I; FALSE NEG.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
Complete proteome.
FT SITE 29 39 "HIGH" REGION.
FT SITE 264 268 "KMSKS" REGION.
FT BINDING 267 267 ATP (By SIMILARITY).
SQ SEQUENCE 467 AA; 54523 MW; B836F5F0CD82BF6B CRC64;

Query Match      1.9%; Score 7; DB 1; Length 467;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 280 IRLLEE 286
    |||||

```

```

Db 421 IRLLEE 427

RESULT 75
K1C0_XENLA
ID K1C0_XENLA      STANDARD;      PRT;      486 AA.
AC P02537;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Keratin 3, type I cytoskeletal 51 kDa (51 kDa cyokeratin).
OS Xenopus laevis (African clawed frog).
OC Sauriata; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89005622; PubMed=2458965;
RA Hoffmann W., Sterrer S., Koenigstorfer A.;
RT "Amino acid sequence microheterogeneities of a type I cyokeratin of
Mr 51,000 from Xenopus laevis epidermis.";
RL FEBS Lett. 237:178-182(1988).
RN [2]
RP SEQUENCE OF 339-386 FROM N.A.
RX MEDLINE=84261417; PubMed=6204859;
RA Hoffmann W., Franz J.K.;
RT "Amino acid sequence of the carboxy-terminal part of an acidic type I
cyokeratin of molecular weight 51 000 from Xenopus laevis epidermis
as predicted from the cDNA sequence.";
RL EMBL J. 3:1301-1306(1984).
CC -!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC -!- MISCELLANEOUS: There are two types of cytoskeletal and
microfibrillar keratin, I (acidic) and II (neutral to basic) (40-
55 and 56-70 kDa, respectively).
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Y00968; CAA68783.1; -.
DR PIR; S01327; KEXL.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin.
FT DOMAIN 1 125 HEAD.
FT DOMAIN 126 438 ROD.
FT DOMAIN 439 486 TAIL.
FT DOMAIN 126 161 COIL 1A.
FT DOMAIN 162 184 LINKER 1.
FT DOMAIN 185 276 COIL 1B.
FT DOMAIN 277 299 LINKER 12.
FT DOMAIN 300 438 COIL 2.
FT SITE 380 380 STUTTER.
FT CONFLICT 352 352 R -> T (IN REF. 2).
SQ SEQUENCE 486 AA; 51889 MW; 9B6D026C597109C5 CRC64;

Query Match      1.9%; Score 7; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241
    |||||
Db 70 GGGAGGG 76

```

Search completed: December 17, 2003, 06:33:52  
Job time : 14.4735 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:26:06 ; Search time 36.1811 Seconds  
(without alignments)  
2624.667 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 388

Sequence: 1 MKLTVTKLKGTHFEIRVQPN.....CDRNEELAANYLLEHAGED 368

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

SPTREMBL 23.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	5.7	365	10 Q94CE9	Q94ce9 arabidopsis
2	21	5.7	367	10 Q9MA10	Q9ma10 arabidopsis
3	21	5.7	371	10 Q8LA46	Q8la46 arabidopsis
4	15	4.1	113	10 Q9SA20	Q9sa20 arabidopsis
5	15	4.1	378	10 Q94C35	Q94c35 arabidopsis
6	15	4.1	378	10 Q9FF16	Q9ff16 arabidopsis
7	15	4.1	379	10 Q03991	Q03991 daucus caro
8	15	4.1	382	10 Q03990	Q03990 daucus caro
9	14	3.8	246	10 Q9S9L8	Q9s9l8 arabidopsis
10	14	3.8	389	10 Q9STA6	Q9sta6 lycopersico
11	13	3.5	392	10 Q40742	Q40742 cryza sativ
12	11	3.0	343	5 Q8IMB7	Q8imb7 drosophila
13	11	3.0	414	5 Q9XZ50	Q9xze0 drosophila
14	11	3.0	414	5 Q9V3W9	Q9v3w9 drosophila
15	11	3.0	419	10 Q9M887	Q9m887 arabidopsis
16	10	2.7	65	10 Q9SCA8	Q9sca8 lycopersico

17	9	2.4	225	16	Q9RU37	Q9ru37 deinococcus
18	9	2.4	753	10	Q9C552	Q9c552 arabidopsis
19	8	2.2	111	10	Q39682	Q39682 daucus caro
20	8	2.2	125	11	Q8CA15	Q8ca15 mus musculus
21	8	2.2	139	10	Q9FY9	Q9fy9 nicotiana t
22	8	2.2	167	11	Q924J3	Q924j3 tamias stri
23	8	2.2	173	10	Q41191	Q41191 arabidopsis
24	8	2.2	174	10	Q9LTP5	Q9ltp5 arabidopsis
25	8	2.2	186	3	Q96V99	Q96v99 pneumocysti
26	8	2.2	202	10	Q9LGP0	Q9lgp0 oryza sativ
27	8	2.2	208	10	Q8LA18	Q8la18 oryza sativ
28	8	2.2	210	10	Q41187	Q41187 arabidopsis
29	8	2.2	220	10	Q39367	Q39367 brassica ol
30	8	2.2	228	11	Q8C4G2	Q8c4g2 mus musculus
31	8	2.2	228	11	Q8BPM6	Q8bpm6 mus musculus
32	8	2.2	307	10	Q8S7T7	Q8s7t7 oryza sativ
33	8	2.2	340	16	Q9KYD8	Q9kyd8 streptomyce
34	8	2.2	360	11	Q9Z2C8	Q9z2c8 mus musculus
35	8	2.2	366	6	Q9GLB8	Q9glb8 cephalorhyn
36	8	2.2	373	3	Q9P639	Q9p639 neurospora
37	8	2.2	384	11	Q9ROJ6	Q9roj6 apodemus sp
38	8	2.2	384	11	Q9ROH6	Q9roh6 rattus norv
39	8	2.2	384	11	Q9ROJ5	Q9roj5 apodemus pe
40	8	2.2	384	11	Q9ROJ4	Q9roj4 apodemus ag
41	8	2.2	384	11	Q9ROH7	Q9roh7 tokudaia os
42	8	2.2	387	10	Q8GVV7	Q8gvv7 oryza sativ
43	8	2.2	390	10	Q9M3Y2	Q9m3y2 triticum ae
44	8	2.2	395	11	Q64217	Q64217 spalax zemn
45	8	2.2	400	10	Q8GTL0	Q8gtl0 oryza sativ
46	8	2.2	401	10	Q9C518	Q9c518 arabidopsis
47	8	2.2	420	10	Q8H010	Q8h010 oryza sativ
48	8	2.2	423	11	Q8KIQ2	Q8kiq2 rattus norv
49	8	2.2	424	11	Q8K1P1	Q8k1p1 tachyorycte
50	8	2.2	427	5	Q9GR85	Q9gr85 halocynthia
51	8	2.2	438	5	Q9NG13	Q9ng13 branchiost
52	8	2.2	458	16	Q8EJY7	Q8ejy7 xanthomonas
53	8	2.2	458	16	Q8P8G0	Q8p8g0 xanthomonas
54	8	2.2	492	4	Q9UNW9	Q9unw9 homo sapien
55	8	2.2	492	11	Q35392	Q35392 mus musculus
56	8	2.2	498	4	Q43267	Q43267 homo sapien
57	8	2.2	544	5	Q8I6C7	Q8i6c7 drosophila
58	8	2.2	544	5	Q8I6C6	Q8i6c6 drosophila
59	8	2.2	544	5	Q8I6C5	Q8i6c5 drosophila
60	8	2.2	544	5	Q8I6C4	Q8i6c4 drosophila
61	8	2.2	544	5	Q8I6C3	Q8i6c3 drosophila
62	8	2.2	544	5	Q8I6B7	Q8i6b7 drosophila
63	8	2.2	544	5	Q8I072	Q8i072 drosophila
64	8	2.2	548	5	Q22809	Q22809 caenorhabdi
65	8	2.2	550	5	Q8IGB4	Q8igb4 drosophila
66	8	2.2	566	5	Q9VTJ7	Q9vtj7 drosophila
67	8	2.2	579	10	Q8LRM4	Q8lrm4 chlamydomon
68	8	2.2	579	10	Q8LRM5	Q8lrm5 chlamydomon
69	8	2.2	584	16	Q9FCJ3	Q9fcj3 streptomyce
70	8	2.2	712	5	Q43981	Q43981 elmeria ten
71	8	2.2	731	10	Q9AWU2	Q9awu2 oryza sativ
72	8	2.2	738	5	O02402	O02402 pinctada fu
73	8	2.2	795	5	O62006	O62006 branchiost
74	8	2.2	809	10	Q43715	Q43715 pisum sativ
75	8	2.2	823	5	Q25343	Q25343 leishmania
76	8	2.2	968	5	Q9VKL0	Q9vkl0 drosophila
77	8	2.2	1004	16	Q8NP40	Q8np40 corynebacte
78	8	2.2	1010	2	Q45340	Q45340 bordetella
79	8	2.2	1014	5	Q9V9D9	Q9v9d9 drosophila
80	8	2.2	1234	11	Q8VD34	Q8vd34 mus musculus
81	8	2.2	1353	5	Q9V9Q7	Q9v9q7 drosophila
82	8	2.2	1385	5	Q8T0L7	Q8t0l7 drosophila
83	8	2.2	1385	5	Q9VVC7	Q9vvc7 drosophila
84	8	2.2	1457	5	Q8SY47	Q8sy47 drosophila
85	8	2.2	1718	5	O62603	O62603 trypanosoma
86	8	2.2	2075	5	Q9VXY2	Q9vxy2 drosophila
87	8	2.2	2425	17	O28859	O28859 archaeoglob
88	8	2.2	2904	11	Q9EPN0	Q9epn0 mus musculus
89	8	2.2	2931	11	Q9EPM9	Q9epm9 mus musculus

90 8 2.2 2936 11 Q9EPN1 Q9epn1 mus musculus  
 91 8 2.2 3016 16 P73590 P73590 synchocyst  
 92 7 1.9 37 6 Q9TRD7 Q9trd7 oryctolagus  
 93 7 1.9 42 10 Q23471 Q23471 arabidopsis  
 94 7 1.9 47 10 Q9S857 Q9s857 glycine max  
 95 7 1.9 49 11 Q9WUR8 Q9wur8 rattus norv  
 96 7 1.9 68 4 Q9UKI4 Q9uki4 homo sapien  
 97 7 1.9 71 2 Q93KA3 Q93ka3 erwinia chr  
 98 7 1.9 72 2 Q50370 Q50370 mycoplasma  
 99 7 1.9 80 5 Q8MKR0 Q8mkr0 drosophila  
 100 7 1.9 80 10 Q8H778 Q8h778 arabidopsis

## ALIGNMENTS

RESULT 1  
 ID Q94CE9 PRELIMINARY; PRT; 365 AA.  
 AC Q94CE9;  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Putative RAD23 protein.  
 GN F20B17.8.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,  
 RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,  
 RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,  
 RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,  
 RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,  
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full Length cDNA of gene F20B17.8 (GI:7715605).";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Banno P., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,  
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,  
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,  
 RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Full Length cDNA of gene F20B17.8 (GI:7715605).";  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY034912; AAK59419.1; -  
 DR EMBL; AY063103; AAL34277.1; -  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR006636; ST11.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR SMART; SM00727; ST11; 1.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; P850053; UBIQUITIN\_2; 1.  
 SQ SEQUENCE 365 AA; 39157 MW; 4FED9EC59B467745 CRC64;

Query Match 5.7%; Score 21; DB 10; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-12;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 VIEAPLACDRNEELAANYLLE 362

Db 338 VIEAPLACDRNEELAANYLLE 358

## RESULT 2

ID Q9MA10 PRELIMINARY; PRT; 367 AA.  
 AC Q9MA10;

DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE F20B17.8.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]

RP SEQUENCE FROM N.A.

RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,  
 RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
 RA Ecker J.R.;

RT "Genomic sequence for Arabidopsis thaliana BAC F20B17 from chromosome

1.,"

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,  
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
 RA Theologis A., Ecker J.;

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC010793; AAF68123.1; -

DR HSSP; P54725; 1DV0.

DR InterPro; IPR004806; Rad23.

DR InterPro; IPR000449; UBA domain.

DR InterPro; IPR000626; Ubiquitin.

DR Pfam; PF00627; UBA; 2.

DR Pfam; PF00240; ubiquitin; 1.

DR SMART; SM00165; UBA; 2.

DR SMART; SM00213; UBQ; 1.

DR TIGRFAMs; TIGR00601; rad23; 1.

DR PROSITE; P850053; UBIQUITIN\_2; 1.

SQ SEQUENCE 367 AA; 39790 MW; C6D01134B5C69CE6 CRC64;

Query Match 5.7%; Score 21; DB 10; Length 367;

Best Local Similarity 100.0%; Pred. No. 2.5e-12;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 VIEAPLACDRNEELAANYLLE 362

Db 340 VIEAPLACDRNEELAANYLLE 360

## RESULT 3

ID Q8LA46

AC Q8LA46 PRELIMINARY; PRT; 371 AA.

QY 342 VIEAPLACDRNEELAANYLLE 362

```

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE DNA repair protein RAD23, putative.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP
RA Haas B.J., Volfovsky N., Town C.D., Troupkhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0(2002).
[2]
RN SEQUENCE FROM N.A.
RP
RA Brover V., Troupkhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY088037; AA065503.1; -.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; STI1.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR SMART; SM00727; STI1; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS0053; UBIQUITIN 2; 1.
SQ SEQUENCE 371 AA; 39747 MW; 081493086EA976E7 CRC64;

Query Match 5.7%; Score 21; DB 10; Length 371;
Best Local Similarity 100.0%; Pred. No. 2.5e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 VIEAFLACDRNEELAANYLLE 362
DB 344 VIEAFLACDRNEELAANYLLE 364

RESULT 4
Q9SA20 ID Q9SA20 PRELIMINARY; PRT; 113 AA.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE F309.1 protein.
GN F309.1
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP
RA Vysotskaia V.S., Schwartz J., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J., Li J., Krenetskaia I., Liu A., Lueros J., Gonzalez A.,
RA Altafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
RA Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "The sequence of BAC F309 from Arabidopsis thaliana chromosome 1.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP
RA STRAIN=cv. Columbia;
RA Vysotskaia V.S., Schwartz J., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J., Li J., Krenetskaia I., Liu A., Lueros J., Gonzalez A.,
RA Altafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
RA Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "The sequence of BAC F309 from Arabidopsis thaliana chromosome 1.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; AC006341; AAD34676.1; -.
DR HSSP; P54725; IDV0.
DR InterPro; IPR000449; UBA_domain.
DR Pfam; PF00627; UBA; 1.
DR SMART; SM00165; UBA; 1.
SQ SEQUENCE 113 AA; 12947 MW; 8BE7CB15C9E12380 CRC64;

Query Match 4.1%; Score 15; DB 10; Length 113;
Best Local Similarity 100.0%; Pred. No. 9.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 CORNEELAANYLLEH 363
DB 93 CORNEELAANYLLEH 107

RESULT 5
Q94C35 ID Q94C35 PRELIMINARY; PRT; 378 AA.
AC Q94C35;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 40.1 kDa protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Shinn P.,
RA Tracy S.E., Banh J., Bowser L., Carninci P., Chung M.K.,
RA Goldsmith A.D., Hayaishizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Pham P.K., Quach H.L.,
RA Sakano H., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C.,
RA Toriumi M., Yamada K., Yu G., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY037181; AAK59766.1; -.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR000636; STI1.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00727; STI1; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS0053; UBIQUITIN_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 378 AA; 40075 MW; E49B591527A3EC27 CRC64;

Query Match 4.1%; Score 15; DB 10; Length 378;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 NNPERAVEYLVSIP 188
DB 177 NNPERAVEYLVSIP 191

RESULT 6
Q9FF16 ID Q9FF16 PRELIMINARY; PRT; 378 AA.
AC Q9FF16;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE DNA repair protein RAD23 homolog (Hypothetical protein)

```

DE (At5g38470/At5g38470).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
OC eucotids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Columbia;  
RC MEDLINE=97471969; PubMed=9330910;  
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
RA Miyajima N., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
RT features of the 1.6 Mb regions covered by twenty physically assigned  
RT pl. clones";  
RL DNA Res. 4:215-230 (1997).  
[2]  
RN SEQUENCE FROM N.A.  
RP Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT "Arabidopsis cDNA clones";  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RP Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,  
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,  
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,  
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,  
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,  
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT "Arabidopsis ORF clones";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
[4]  
RN SEQUENCE FROM N.A.  
RP Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,  
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length messenger RNA sequences greatly improve genome  
RT annotation";  
RL Genome Biol. 0:0-0 (2002).  
[5]  
RN SEQUENCE FROM N.A.  
RP Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.;  
RT "Full-length cDNA from Arabidopsis thaliana";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB05248; BAB09359.1; -;  
DR EMBL; AY058196; AAL25609.1; -;  
DR EMBL; AY081835; AAL87405.1; -;  
DR EMBL; AY087564; AAL65106.1; -;  
DR HSSP; P54725; IDVO.  
DR InterPro; IPR004806; Rad23.  
DR InterPro; IPR000449; UBA domain.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00627; UBA; 2.  
DR TIGRfam; TIGR00601; rad23; 1.  
DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 378 AA; 40066 MW; DD9B59153EA3B6A9 CRC64;

Query Match 4.1%; Score 15; DB 10; Length 378;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 174 NNPERAVEVLYSGIP 188

Db 177 NNPERAVEVLYSGIP 191  
|||||  
RESULT 7  
O03991 PRELIMINARY; PRT; 379 AA.  
ID O03991  
AC O03991; (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE RAD23 protein, isoform II.  
OS Daucus carota (Carrot).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;  
OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.  
OX NCBI\_TaxID=4039;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=W001C;  
RC MEDLINE=98345997; PubMed=9681019;  
RA Sturm A., Leinhardt S.;  
RT "Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in  
RT yeast";  
RL Plant J. 13:815-821 (1998).  
DR EMBL; Y12014; CAA72742.1; -;  
DR HSSP; P54725; IDVO.  
DR InterPro; IPR004806; Rad23.  
DR InterPro; IPR006636; STIL.  
DR InterPro; IPR000449; UBA domain.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00627; UBA; 2.  
DR SMART; SM00727; STIL; 1.  
DR SMART; SM00165; UBA; 2.  
DR SMART; SM00213; UBQ; 1.  
DR TIGRfam; TIGR00601; rad23; 1.  
DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
SQ SEQUENCE 379 AA; 40530 MW; B26697B439CC5929 CRC64;  
Query Match 4.1%; Score 15; DB 10; Length 379;  
Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 342 VIEAFLACDRNEELA 356  
|||||  
Db 352 VIEAFLACDRNEELA 366  
|||||  
RESULT 8  
O03990 PRELIMINARY; PRT; 382 AA.  
ID O03990  
AC O03990; (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE RAD23, isoform I.  
OS Daucus carota (Carrot).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;  
OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.  
OX NCBI\_TaxID=4039;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=W001C;  
RC MEDLINE=98345997; PubMed=9681019;  
RA Sturm A., Leinhardt S.;  
RT "Two isoforms of plant RAD23 complement a UV-sensitive rad23 mutant in  
RT yeast";  
RL Plant J. 13:815-821 (1998).  
DR EMBL; Y12013; CAA72741.1; -;  
DR HSSP; P54725; IDVO.  
DR InterPro; IPR002965; P\_rich\_extens.

DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR006636; Still.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR SMART; SM00727; Still; 1.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS0053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 382 AA; 40344 MW; 27FF02A2402F3DC2 CRC64;

Query Match 4.1%; Score 15; DB 10; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 NNPERAVELYSGIP 188

Db 187 NNPERAVELYSGIP 201

# RESULT 9

ID Q9S9L8 PRELIMINARY; PRT; 246 AA.  
 AC Q9S9L8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE T24D18.27 protein.  
 GN T24D18.27.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 [1]

SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Liu S.X., Yu G., Sakano H., Jhaveri A., Lee J.M., Lenz C., Pham P.,  
 RA Toriumi M., Chin C., Chlou J., Choi E., Chung M., Gonzalez A.,  
 RA Hwang B., Koo T., Li J., Liu A., Vayberg M., Altafi H., Brooks S.,  
 RA Buehler E., Chao Q., Conn L., Conway A., Hansen N., Johnson-Hopson C.,  
 RA Khan S., Kim C., Lam B., Nguyen M., Palm C., Shinn P., Tambunga G.,  
 RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;  
 RT "The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 [2]

SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Theologis A.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC010924; AAF18513.1; -.

DR HSSP; P02248; LUBI.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 1.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR SMART; SM00165; UBA; 1.  
 DR SMART; SM00213; UBQ; 1.  
 DR PROSITE; PS0053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 246 AA; 25966 MW; 0867D1E7D72B6FFE CRC64;

Query Match 3.8%; Score 14; DB 10; Length 246;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 167 RALRAAYNNPERAV 180

Db 155 RALRAAYNNPERAV 168

## RESULT 10

Q9STA6 PRELIMINARY; PRT; 389 AA.  
 AC Q9STA6;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE RAD23 protein.  
 GN RAD23.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; famids; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4081;  
 RN Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 [1]

SEQUENCE FROM N.A.  
 RC STRAIN=cv. West Virginia 106; TISSUE=Fruit;  
 RA Lemaire-Chanley M., Petit J., Raymond P., Chevallier C.;  
 RT "Analysis of gene expression during early tomato fruit development by  
 RT mRNA differential display.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ243875; CAB51544.1; -.  
 DR HSSP; P54725; IDV0  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR006636; Still.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR SMART; SM00727; Still; 1.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS0053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 389 AA; 41508 MW; D2BAEDF0FE70778A CRC64;

Query Match 3.8%; Score 14; DB 10; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 167 RALRAAYNNPERAV 180

Db 177 RALRAAYNNPERAV 190

## RESULT 11

ID Q40742 PRELIMINARY; PRT; 392 AA.  
 AC Q40742;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE ORAD23.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartidae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 [1]

SEQUENCE FROM N.A.  
 RC STRAIN=Nipponbare;  
 RX MEDLINE=97369378; PubMed=9225866;  
 RA Schultz T.F., Quatrano R.S.;  
 RT "Characterization and expression of a rice RAD23 gene.";  
 RL Plant Mol. Biol. 34:557-562(1997).  
 DR EMBL; U63530; RAB65841.1; -.

DR HSSP; P54725; IDV0.  
 DR Gramine; Q40742; -.  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR006636; Still.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.

DR Pfam; PF00240; ubiquitous; 1.  
 DR SMART; SM00727; STI1; 1.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; URQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS5053; UBQUITIN\_2; 1.  
 SQ SEQUENCE 392 AA; 41754 MW; BD6E08574CC7CABC CRC64;

Query Match 3.5%; Score 13; DB 10; Length 392;  
 Best Local Similarity 100.0%; Pred. No. 0.00034;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 NNPERAVEYLVS 186  
 |||||  
 Db 187 NNPERAVEYLVS 199

RESULT 12

Q81MB7 Q81MB7 PRELIMINARY; PRT; 343 AA.  
 AC Q81MB7  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE CG1836-PB.  
 GN RAD23.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celniker S.E., Scher S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Chalker S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers V.H., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Gabor G.L.,  
 RA April J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreenek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzales M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Paragase V., Park S., Patel S., Pfeiffer B., Scheeler F.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003844; AAN06526.1; -;  
 SQ SEQUENCE 343 AA; 37549 MW; 2DFC676B41852079 CRC64;

Query Match 3.0%; Score 11; DB 5; Length 343;  
 Best Local Similarity 100.0%; Pred. No. 0.031;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 YNNPERAVEYL 183  
 |||||  
 Db 113 YNNPERAVEYL 123

RESULT 13

Q9XZE0 Q9XZE0 PRELIMINARY; PRT; 414 AA.  
 AC Q9XZE0;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE DHR23.  
 OS RAD23 OR DHR23 OR CG1836.  
 GN Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nabrochkin E.N., Grischuk A.L., Soldatov A.V.;  
 RT "Cloning and characterization of the Drosophila melanogaster homologue  
 of the Saccharomyces cerevisiae gene RAD23";  
 RL Genetika 35:0-0(1999).  
 DR EMBL; AF136506; RAD33695.1; -;  
 DR HSSP; P54725; LDV0.  
 DR FlyBase; FBgn0026777; Rad23.  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR006636; STI1.  
 DR InterPro; IPR000449; UBA\_domain.



DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR SMART; SM00727; Ubiquitin; 1.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS0053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 414 AA; 45794 MW; 4C3FF94CAll16F7AB CRC64;

Query Match 3.0%; Score 11; DB 5; Length 414;  
 Best Local Similarity 100.0%; Pred.No. 0.038; 0; Indels 0; Gaps 0;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 YNNPERAVEYL 183  
 Db 184 YNNPERAVEYL 194

## RESULT 14

Q9V3W9  
 ID Q9V3W9 PRELIMINARY; PRT; 414 AA.  
 AC Q9V3W9  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE RAD23 protein.  
 GN RAD23 OR CG1836.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA LaRoche P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murthy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Sheng H.,  
 RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.

RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Brodsky M.H., Rubin G.M., Tsang G.;  
 RT "Full length Drosophila melanogaster cDNA sequence.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB003844; AAF59352.1; -  
 DR EMBL; AF132147; AAD33594.1; -  
 DR HSP; P54725; 1DV0.  
 DR FlyBase; FBgn0026777; Rad23.  
 DR InterPro; IPR004806; Rad23.  
 DR SMART; SM00636; STIL.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR SMART; SM00727; STIL; 1.  
 DR SMART; SM00165; UBA; 2.  
 DR SMART; SM00213; UBQ; 1.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PS0053; UBIQUITIN 2; 1.  
 SQ SEQUENCE 414 AA; 45780 MW; 4C2E494CAll16F7AB CRC64;

Query Match 3.0%; Score 11; DB 5; Length 414;  
 Best Local Similarity 100.0%; Pred.No. 0.038; 0; Indels 0; Gaps 0;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 YNNPERAVEYL 183  
 Db 184 YNNPERAVEYL 194

## RESULT 15

Q9M887  
 ID Q9M887 PRELIMINARY; PRT; 419 AA.  
 AC Q9M887;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative RAD23 (AT3G02540/F16B3\_17).  
 GN F16B3.17.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,  
 RA Renning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
 RT "Arabidopsis thaliana chromosome III BAC F16B3 genomic sequence.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Gibson H.A., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,  
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,  
 RA Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis cDNA clones.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Shinn P., Banh J., Bowser L.,  
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,

RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis ORF clones";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC021640; AAF32461.1; -;  
 DR EMBL; AY039562; AAK62617.1; -;  
 DR EMBL; AY113034; AAM47342.1; -;  
 DR HSSP; P54725; IDV0.  
 DR InterPro; IPR002965; P rich\_extensin.  
 DR InterPro; IPR004806; Rad23.  
 DR InterPro; IPR000449; UBA domain.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00627; UBA; 2.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR TIGRFAMs; TIGR00601; rad23; 1.  
 DR PROSITE; PSS0053; UBIQUITIN\_2; 1.  
 SQ SEQUENCE 419 AA; 44247 MW; CAA13BC4FFEB1E25 CRC64;

Query Match 3.0%; Score 11; DB 10; Length 419;  
 Best Local Similarity 100.0%; Pred. No. 0.038;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VKTLKGTHFEI 15  
 |||||  
 Db 5 VKTLKGTHFEI 15

RESULT 16  
 Q9SCAB PRELIMINARY; PRT; 65 AA.  
 AC Q9SCAB;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Rad23 protein (Fragment).  
 GN RAD23.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. West virginia 106; TISSUE=Fruit;  
 RA Lemaire-Chamley M., Petit J., Causee M., Raymond P., Chevalier C.;  
 RT "Isolation and characterization of cDNAs expressed during early  
 development of tomato fruit by mRNA differential display."  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ270958; CAB65692.1; -;  
 DR HSSP; P54725; IDV0.  
 DR InterPro; IPR000449; UBA\_domain.  
 DR Pfam; PF00627; UBA; 1.  
 DR SMART; SM00165; UBA; 1.  
 FT NON TER 1 1  
 SQ SEQUENCE 65 AA; 7233 MW; 0538945EE9030B4D CRC64;

Query Match 2.7%; Score 10; DB 10; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 0.065;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 NEELAANYLL 361  
 |||||  
 Db 48 NEELAANYLL 57

RESULT 17  
 Q9RU37 PRELIMINARY; PRT; 225 AA.  
 AC Q9RU37;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE DNA-binding response regulator.  
 GN DR1558.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
 OC Deinococcaceae; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Panghile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER  
 CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.  
 CC -1- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC EMBL; AE001999; AAF11120.1; -;  
 DR HSSP; P10957; 1RNL.  
 DR TIGR; DR1558; -;  
 DR InterPro; IPR000792; HTH\_LuxR.  
 DR InterPro; IPR001789; Response\_reg.  
 DR Pfam; PF00196; GerE; 1.  
 DR Pfam; PF00072; Response\_reg; 1.  
 DR PRINTS; PR00038; HTHLUXR.  
 DR ProDom; PD000307; HTH\_LuxR; 1.  
 DR ProDom; PD000039; Response\_reg; 1.  
 DR SMART; SM00421; HTH\_LUXR; 1.  
 DR SMART; SM00448; REC; 1.  
 DR PROSITE; PS0110; RESPONSE\_REGULATORY; 1.  
 DR PROSITE; PS0110; RESPONSE\_REGULATORY; 1.  
 KW DNA-binding; Phosphorylation; Sensory transduction;  
 KW Transcription regulation; Complete proteome.  
 SQ SEQUENCE 225 AA; 24452 MW; 953B45742CB2A152 CRC64;

Query Match 2.4%; Score 9; DB 16; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 STLEENKVN 65  
 |||||  
 Db 95 STLEENKVN 103

RESULT 18  
 Q9C552 PRELIMINARY; PRT; 753 AA.  
 ID Q9C552;  
 AC Q9C552;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Putative RNA-binding domain (Hypothetical 84.8 kDa protein) (Putative  
 RNA-binding protein).  
 GN T10D10.21 OR T9N14.7.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,

RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
RA Dunn P., Eguu P., Feldbiyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,  
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
RA Militischer J., Miranda M., Nguyen M., Nierman M.C., Osborne B.I.,  
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
RA Utterback T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,  
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
RT thaliana.";  
RL Nature 408:816-820(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,  
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,  
RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koseena E., Lam B.,  
RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,  
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
RT "Full Length cDNA of gene T10D10.21 (GI:12325275).";  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,  
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,  
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koseena E.,  
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,  
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,  
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
RT "Full Length cDNA of gene T10D10.21 (GI:12325275).";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC016529; AAG52582.1; -;  
DR EMBL; AC067754; AAG51785.1; -;  
DR EMBL; AY051073; AAK93750.1; -;  
DR EMBL; AY035041; AAK95946.1; -;  
DR InterPro; IPR001313; Pumilio/Puf.  
DR Pfam; PF00806; PUF; 5.  
DR SMART; SM00025; Pumilio; 5.  
KW Hypothetical protein.  
SQ SEQUENCE 753 AA; 84827 MW; 300CC905280C29CA CRC64;

Query Match 2.4%; Score 9; DB 10; Length 753;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 340 ARVIEAFLA 348  
Db 542 ARVIEAFLA 550  
|||||

RESULT 19  
Q39682  
ID Q39682 PRELIMINARY; PRT; 111 AA.  
AC Q39682  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Glycine-rich protein (Fragment).  
OS Daucus carota (Carrot).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.  
OX NCBI\_TaxID=4039;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=Danver Half-long; TISSUE=Somatic embryos at the globular stage;  
RA Lin X., Hwang G.-J., Zimmerman J.L.;  
RT "Isolation and characterization of a diverse set of genes from carrot  
RT somatic embryos";  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U47097; A801097.1; -;  
FT NON TER 1  
SQ SEQUENCE 111 AA; 13829 MW; 082789CCADF1C2A CRC64;

Query Match 2.2%; Score 8; DB 10; Length 111;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
Db 18 AGGGAGGG 25  
|||||

RESULT 20  
Q8CA15  
ID Q8CA15 PRELIMINARY; PRT; 125 AA.  
AC Q8CA15  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Hypothalamus;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK038724; BAC30110.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 125 AA; 13205 MW; 58AD04310B540101 CRC64;

Query Match 2.2%; Score 8; DB 11; Length 125;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
Db 82 AGGGAGGG 89  
|||||

RESULT 21  
Q9FOY9  
ID Q9FOY9 PRELIMINARY; PRT; 139 AA.  
AC Q9FOY9  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE AVR9/Cf-9 rapidly elicited protein 75.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Petite Havana;  
RA Durrant W.E., Rowland O., Piedras P., Hammond-Kosack K.E.,  
RA Jones J.D.G.;  
RT "cDNA expression profiling reveals rapid, resistance gene-dependent,  
RT active oxygen-independent, gene induction during the plant defense

RT response.";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF211540; AAG43558.1; -.  
 SQ SEQUENCE 139 AA; 14597 MW; 1D485BD90F811D9C CRC64;

Query Match 2.2%; Score 8; DB 10; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
 Db 87 AGGGAGGG 94  
 |||||

## RESULT 22

Q924J3 ID Q924J3 PRELIMINARY; PRT; 167 AA.  
 AC Q924J3  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Interphotoreceptor binding protein (Fragment).  
 GN IRBP.  
 OS Tamas striatus (Eastern chipmunk).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Scuriinae;  
 OC Tamas.  
 OC NCBI\_TaxID=45474;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21240538; PubMed=11341810;  
 RA DeBry R.W., Sagel R.M.;  
 RT "Phylogeny of Rodentia (Mammalia) inferred from the nuclear-encoded  
 RT gene IRBP";  
 RL Mol. Phylogenet. Evol. 19:290-301(2001).  
 DR EMBL; AF297282; AAK62261.1; -.  
 DR InterPro; IPR003581; TSPC.  
 DR Pfam; PF02692; IRBP; 1.  
 KW Receptor.  
 FT NON TER 1  
 FT NON TER 167  
 FT NON TER 167  
 SQ SEQUENCE 167 AA; 18463 MW; 0ED660DB40D94819 CRC64;

Query Match 2.2%; Score 8; DB 11; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 LEAPQQAP 109  
 Db 41 LEAPQQAP 48  
 |||||

## RESULT 23

Q41191 ID Q41191 PRELIMINARY; PRT; 173 AA.  
 AC Q41191;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Glycine-rich protein (Fragment).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Arabidopsida; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsida.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93044485; PubMed=2152168;  
 RA De Oliveira D.E., Seurinck J., Inze D., Van Montagu M., Botterman J.;  
 RT "Differential expression of five Arabidopsis genes encoding glycine-  
 RT rich proteins.";  
 RL Plant Cell 2:427-436(1990).  
 DR EMBL; S47414; AAB24077.1; -.

DR InterPro; IPR002952; Eggshell.  
 DR InterPro; IPR000817; Prion.  
 DR PRINTS; PR01228; EGGSHLL.  
 DR PRINTS; PR00341; PRION.  
 FT NON TER 1  
 SQ SEQUENCE 173 AA; 13587 MW; D02259B4B38F80C3 CRC64;

Query Match 2.2%; Score 8; DB 10; Length 173;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
 Db 84 AGGGAGGG 91  
 |||||

## RESULT 24

Q9LTP5 ID Q9LTP5 PRELIMINARY; PRT; 174 AA.  
 AC Q9LTP5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Genomic DNA, chromosome 3, Pl clone: MGC12 (AF392047/MGC12\_23).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsida.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=20277480; PubMed=10819329;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
 RT features of the regions of 4,504,864 bp covered by sixty Pl and TAC  
 RT clones.";  
 RL DNA Res. 7:131-135(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Koesema E., Chen H., Cheuk R., Kim C.J., Meyers M.C., Shinn P.,  
 RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,  
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis ORF clones.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Gibson H.A., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,  
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,  
 RA Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis cDNA clones.";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB024036; BAB02830.1; -.  
 DR EMBL; AY058226; AAL15400.1; -.  
 DR EMBL; AY045629; AAK73987.1; -.  
 DR InterPro; IPR002952; Eggshell.  
 DR InterPro; IPR000817; Prion.  
 DR PRINTS; PR01228; EGGSHLL.

DR PRINTS; PRO0341; PRION.  
SQ SEQUENCE 174 AA; 13718 MW; 196DF768A69EA176 CRC64;

Query Match 2.2%; Score 8; DB 10; Length 174;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
| | | | |  
DB 85 AGGAGGG 92

## RESULT 25

Q96V99 PRELIMINARY; PRT; 186 AA.  
AC Q96V99;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Dihydrofolate reductase (SC 1.5.1.3) (Fragment).  
OS Pneumocystis carinii f. sp. muris.  
OC Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes; Pneumocystidaceae;  
OC Pneumocystis.  
OC NCBI\_TaxID=42066;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21537003; PubMed=11679931;  
RA Ma L., Imachi H., Sukura A., Kovacs J.A.;  
RT "Genetic Divergence of the Dihydrofolate Reductase and Dihydropteroate  
RT Synthase Genes in Pneumocystis carinii from 7 Different Host  
RT Species";  
RL J. Infect. Dis. 184:1358-1362 (2001).  
DR EMBL; AF175561; AAL25644.1; -.  
DR InterPro; IPR001796; DHFR.  
DR Pfam; PF00186; Dihfolate\_red; 1.  
DR PRINTS; PRO0070; DHFR.  
DR PROSITE; PS00075; DHFR; 1.  
KW Oxidoreductase.  
FT NON\_TER 1 186  
FT TER 186 186  
SQ SEQUENCE 186 AA; 21402 MW; 1A181FDSAA5C8600 CRC64;

Query Match 2.2%; Score 8; DB 3; Length 186;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 KVNEDGFL 70  
| | | | |  
DB 178 KVNEDGFL 185

## RESULT 26

Q9LGP0 PRELIMINARY; PRT; 202 AA.  
AC Q9LGP0;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE P0684C01.5 protein.  
GN P0684C01.5.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OC NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
RT P0684C01.1";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002487; BAB07944.1; -.

DR Gramene; Q9LGP0; -.  
SQ SEQUENCE 202 AA; 20969 MW; 44594A6CFE37B7F8 CRC64;

Query Match 2.2%; Score 8; DB 10; Length 202;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 242  
| | | | |  
DB 24 GGGAGGG 31

## RESULT 27

Q8L418 PRELIMINARY; PRT; 208 AA.  
AC Q8L418;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE P0551A11.6 protein (QJ1116 C07.6 protein).  
GN P0551A11.6 OR QJ1116 C07.6  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OC NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
RT clone:P0551A11.1";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC  
RT clone:QJ1116 C07.1";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF003934; BAB92814.1; -.  
DR EMBL; AP004253; BAC00689.1; -.  
DR Gramene; Q8L418; -.  
SQ SEQUENCE 208 AA; 20670 MW; C4D641DBA4D55B80 CRC64;

Query Match 2.2%; Score 8; DB 10; Length 208;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
| | | | |  
DB 71 AGGAGGG 78

## RESULT 28

Q41187 PRELIMINARY; PRT; 210 AA.  
AC Q41187;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Glycine-rich protein (Fragment).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OC NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93044485; PubMed=2152168;  
RA de Oliveira D.E., Seurinck J., Inze D., Van Montagu M., Botterman J.;  
RT "Differential expression of five Arabidopsis genes encoding glycine-  
RT rich proteins.";

```

RL Plant Cell 2:427-436(1990).
DR EMBL; S47405; AAB24073.1; -.
FT NON TER 1
SQ SEQUENCE 210 AA; 14445 MW; 03A5890F258204BC CRC64;

Query Match 2.2%; Score 8; DB 10; Length 210;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241
Db 148 AGGAGGG 155

RESULT 29
Q39367
ID Q39367 PRELIMINARY; PRT; 220 AA.
AC Q39367;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glycine-rich protein (Fragment).
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=alaboglabra; TISSUE=Stigma;
RA Booker J.P.;
RL Thesis (1996), Biological Science, University of Durham.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=alaboglabra; TISSUE=Stigma;
RA Scutt C.P.;
RL Thesis (1990), Biological Science, University of Durham.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=alaboglabra; TISSUE=Stigma;
RA Croy R.D.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; 274892; CAA99171.1; -.
DR InterPro; IPR000817; Prion.
DR PRINTS; PR00341; PRION.
FT NON TER 1
SQ SEQUENCE 220 AA; 15368 MW; 7ED3ED15D717B519 CRC64;

Query Match 2.2%; Score 8; DB 10; Length 220;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241
Db 16 AGGAGGG 23

RESULT 30
Q8C4G2
ID Q8C4G2 PRELIMINARY; PRT; 228 AA.
AC Q8C4G2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RL MEDLINE=22354683; PubMed=12466851;

```

```

RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK082275; BAC38452.1; -.
KW Hypothetical protein.
SQ SEQUENCE 228 AA; 24880 MW; 0BAD53B878442E43 CRC64;

Query Match 2.2%; Score 8; DB 11; Length 228;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241
Db 82 AGGAGGG 89

RESULT 31
Q8BPM6
ID Q8BPM6 PRELIMINARY; PRT; 228 AA.
AC Q8BPM6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eye;
RL MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK053760; BAC35510.1; -.
KW Hypothetical protein.
SQ SEQUENCE 228 AA; 24908 MW; 5A7F826B78472F08 CRC64;

Query Match 2.2%; Score 8; DB 11; Length 228;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241
Db 82 AGGAGGG 89

RESULT 32
Q8S7T7
ID Q8S7T7 PRELIMINARY; PRT; 307 AA.
AC Q8S7T7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 32.5 kDa protein (Fragment).
GN OSJNBAA0091p11.2.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA Tstirin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pai G.,
RA VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,

```

RA Salzberg S.L., White O., Fraser C.M.;  
 RT "Oryza sativa chromosome 3 BAC OSJNBa0091P11 genomic sequence.";  
 RL EMBL; AC073556; AAL84302.1; -.  
 DR Gramene; Q897T7; -.  
 DR InterPro; IPR006873; DUF620.  
 DR Pfam; PF04788; DUF620; 1.  
 DR Hypothetical protein.  
 FT NON TER  
 SQ SEQUENCE 307 AA; 32503 MW; 8FA3FD1BED4BFBF9 CRC64;

Query Match 2.2%; Score 8; DB 10; Length 307;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
 DB 61 AGGAGGG 68  
 |||||

RESULT 33  
 ID QKYD8 PRELIMINARY; PRT; 340 AA.  
 AC QKYD8;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative secreted transglycosylase.  
 GN SCO4132 OR SCD72A.18.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Brown S.P., Harris D.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Denapante D., Eichner A., Cullum J.,  
 RA Redenbach M., Kieiser H.M., Hopwood D.A.;  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieiser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieiser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,  
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939119; CAB92385.1; -.  
 DR InterPro; IPR000664; NLP\_C\_P60.  
 DR InterPro; IPR000189; SLT domain.  
 DR Pfam; PF00877; NLP\_C\_P60; 1.  
 DR Pfam; PF01464; SLT; 1.

KW Complete proteome.  
 SQ SEQUENCE 340 AA; 35682 MW; 516F68207D31962A CRC64;

Query Match 2.2%; Score 8; DB 16; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
 DB 40 AGGAGGG 47  
 |||||

RESULT 34  
 Q9Z2C8 PRELIMINARY; PRT; 360 AA.  
 ID Q9Z2C8;  
 AC Q9Z2C8;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Y-box protein MSY2.  
 GN YBX2 OR MSY2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD-1; TISSUE=Testis;  
 RX MEDLINE=98455458; PubMed=9780336;  
 RA Gu W., Tekur S., Reinbold R., Spig J.J., Choi Y.C., Zheng J.Z.,  
 RA Murray M.T., Hecht N.B.;  
 RT "Mammalian male and female germ cells express a germ cell-specific Y-  
 Box protein, MSY2.";  
 RL Biol. Reprod. 59:1266-1274(1998).  
 CC 1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.  
 DR EMBL; AF073954; AAC98673.1; -.  
 DR HSSP; P41016; 1C90.  
 DR MGD; MGI:1096372; Ybx2.  
 DR InterPro; IPR002059; Cold\_shock.  
 DR Pfam; PF00313; CSD; 1.  
 DR PRINTS; PR00050; COLDSHOCK.  
 DR ProDom; PD000621; Cold\_shock; 1.  
 DR SMART; SM00357; CSP; 1.  
 DR PROSITE; PS00352; COLD SHOCK; 1.  
 SQ SEQUENCE 360 AA; 38270 MW; DFCAAE7F936731BF CRC64;

Query Match 2.2%; Score 8; DB 11; Length 360;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
 DB 40 AGGAGGG 47  
 |||||

RESULT 35  
 Q9GLB8 PRELIMINARY; PRT; 366 AA.  
 ID Q9GLB8;  
 AC Q9GLB8;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
 DE Interphotoreceptor retinoid-binding protein (Fragment).  
 GN IRBP.  
 OS Cephalorhynchus eutropia (Black dolphin).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;  
 OC Cephalorhynchus.  
 OX NCBI\_TaxID=27608;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cassens I., Vicario S., Waddell V.G., Balchowsky H., Van Belle D.,  
 RA Ding W., Fan C., Mohan L., Simoes-Lopes P.C., Bastida R., Meyer A.,

```

RA Stanhope M.J., Milinkovitch M.C.;
RT "Independent adaptation to riverine habitats allowed survival of
RL ancient cetacean lineages.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF304076; AAC25970.1; -.
DR InterPro; IPR003581; TSPC.
DR Pfam; PF02692; IRBP; 1.
DR SMART; SM00245; TSPC; 1.
KW Receptor.
FT NON_TER 1 366
FT NON_TER 366 366
SQ SEQUENCE 366 AA; 39870 MW; 10A620359F3D1B7A CRC64;

Query Match 2.2%; Score 8; DB 6; Length 366;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 LEAPQAP 109
Db 1 LEAPQAP 8

RESULT 36
Q9P639 PRELIMINARY; PRT; 373 AA.
AC Q9P639;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Related to glycine-rich cell wall structural protein.
GN B21J21.090.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Align V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Meves H.W., Mannhaupt G.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL355929; CAB91331.2; -.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR001185; MS_channel.
DR Pfam; PF01741; MscL; 1.
DR PRINTS; PR01228; EGGSHLL.
SQ SEQUENCE 373 AA; 36389 MW; 7546EBE7260ED192 CRC64;

Query Match 2.2%; Score 8; DB 3; Length 373;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241
Db 253 AGGAGGG 260

RESULT 37
Q9R0J6 PRELIMINARY; PRT; 384 AA.
AC Q9R0J6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Interphotoreceptor retinoid binding protein (Fragment).
GN IRBP.
OS Apodemus speciosus (large Japanese field mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
OC Apodemus.

```

```

OX NCBI_TaxID=105296;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=HS240;
RX MEDLINE=20319711; PubMed=10862357;
RA Serizawa K., Suzuki H., Tsuchiya K.;
RT "A phylogenetic view on species radiation in Apodemus inferred from
RL variation of nuclear and mitochondrial genes.";
RL Biochem. Genet. 38:27-40(2000).
DR EMBL; AB032856; BAA85242.1; -.
DR InterPro; IPR003581; TSPC.
DR Pfam; PF02692; IRBP; 2.
DR SMART; SM00245; TSPC; 1.
KW Receptor.
FT NON_TER 1 384
FT NON_TER 384 384
SQ SEQUENCE 384 AA; 41642 MW; 8E4F40B292305814 CRC64;

Query Match 2.2%; Score 8; DB 11; Length 384;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 LEAPQAP 109
Db 26 LEAPQAP 33

RESULT 38
Q9R0H6 PRELIMINARY; PRT; 384 AA.
AC Q9R0H6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Interphotoreceptor retinoid binding protein (Fragment).
GN IRBP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Sprague-Dawley;
RX MEDLINE=20319711; PubMed=10862357;
RA Serizawa K., Suzuki H., Tsuchiya K.;
RT "A phylogenetic view on species radiation in Apodemus inferred from
RL variation of nuclear and mitochondrial genes.";
RL Biochem. Genet. 38:27-40(2000).
DR EMBL; AB033714; BAA85627.1; -.
DR InterPro; IPR003581; TSPC.
DR Pfam; PF02692; IRBP; 2.
DR SMART; SM00245; TSPC; 1.
KW Receptor.
FT NON_TER 1 384
FT NON_TER 384 384
SQ SEQUENCE 384 AA; 41790 MW; E4B6C02F02B50189 CRC64;

Query Match 2.2%; Score 8; DB 11; Length 384;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 LEAPQAP 109
Db 26 LEAPQAP 33

RESULT 39
Q9R0J5 PRELIMINARY; PRT; 384 AA.
AC Q9R0J5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

```



DE Interphotoreceptor retinoid binding protein (Fragment).  
 GN IRBP.  
 OS Apodemus peninsulae (Korean field mouse).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 OC Apodemus.  
 OX NCBI\_TaxID=105297;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HS329/KT2772;  
 RX MEDLINE=20319711; PubMed=10862357;  
 RA Serizawa K., Suzuki H., Tsuchiya K.;  
 RT "A phylogenetic view on species radiation in Apodemus inferred from  
 variation of nuclear and mitochondrial genes.";  
 RL Biochem. Genet. 38:27-40(2000).  
 DR EMBL: AB032857; BAA85243.1; -;  
 DR InterPro: IPR003581; TSPC.  
 DR Pfam; PF02692; IRBP; 2.  
 DR SMART; SM00245; TSPC; 1.  
 KW Receptor.  
 FT NON\_TER 1 1  
 FT NON\_TER 384 384  
 FT SEQUENCE 384 AA; 41674 MW; F564540D3973EA2C CRC64;

Query Match 2.2%; Score 8; DB 11; Length 384;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LEAPQAP 109  
 Db 26 LEAPQAP 33  
 |||||

## RESULT 40

ID Q9ROJ4 PRELIMINARY; PRT; 384 AA.  
 AC Q9ROJ4;  
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
 DE Interphotoreceptor retinoid binding protein (Fragment).  
 GN IRBP.  
 OS Apodemus agrarius (Eurasian field mouse).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 OC Apodemus.  
 OX NCBI\_TaxID=39030;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HS1253/KT9271;  
 RX MEDLINE=20319711; PubMed=10862357;  
 RA Serizawa K., Suzuki H., Tsuchiya K.;  
 RT "A phylogenetic view on species radiation in Apodemus inferred from  
 variation of nuclear and mitochondrial genes.";  
 RL Biochem. Genet. 38:27-40(2000).  
 DR EMBL: AB032858; BAA85244.1; -;  
 DR InterPro: IPR003581; TSPC.  
 DR Pfam; PF02692; IRBP; 2.  
 DR SMART; SM00245; TSPC; 1.  
 KW Receptor.  
 FT NON\_TER 1 1  
 FT NON\_TER 384 384  
 FT SEQUENCE 384 AA; 41492 MW; BEB5540C77E4D0F CRC64;

Query Match 2.2%; Score 8; DB 11; Length 384;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LEAPQAP 109  
 Db 26 LEAPQAP 33  
 |||||

RESULT 41  
 Q9ROH7  
 ID Q9ROH7 PRELIMINARY; PRT; 384 AA.  
 AC Q9ROH7;  
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
 DE Interphotoreceptor retinoid binding protein (Fragment).  
 GN IRBP.  
 OS Tokudaia osimensis (Ryukyu spiny rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 OC Tokudaia.  
 OX NCBI\_TaxID=73109;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HS1162;  
 RX MEDLINE=20319711; PubMed=10862357;  
 RA Serizawa K., Suzuki H., Tsuchiya K.;  
 RT "A phylogenetic view on species radiation in Apodemus inferred from  
 variation of nuclear and mitochondrial genes.";  
 RL Biochem. Genet. 38:27-40(2000).  
 DR EMBL: AB033712; BAA85873.1; -;  
 DR InterPro: IPR003581; TSPC.  
 DR Pfam; PF02692; IRBP; 2.  
 DR SMART; SM00245; TSPC; 1.  
 KW Receptor.  
 FT NON\_TER 1 1  
 FT NON\_TER 384 384  
 FT SEQUENCE 384 AA; 41695 MW; AD57C8D5175F5475 CRC64;

Query Match 2.2%; Score 8; DB 11; Length 384;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LEAPQAP 109  
 Db 26 LEAPQAP 33  
 |||||

## RESULT 42

ID Q8GVV7 PRELIMINARY; PRT; 387 AA.  
 AC Q8GVV7;  
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE OJ1705\_C03.15 protein.  
 GN OJ1705\_C03.15.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC  
 clone: OJ1705\_C03.15";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF003962; BAC45097.1; -;  
 SQ SEQUENCE 387 AA; 40202 MW; 8426D5490DFB09FB CRC64;

Query Match 2.2%; Score 8; DB 10; Length 387;

Best Local Similarity 100.0%; Pred. No. 39; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
Db 40 AGGGAGGG 47

RESULT 43  
Q9M3Y2  
ID Q9M3Y2 PRELIMINARY; PRT; 390 AA.  
AC Q9M3Y2;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Glycine-rich protein precursor.  
GN GRPL  
OS Triticum aestivum (Wheat).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
OC Triticeae; Triticum.  
OX NCBI\_TaxID=4565;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21107239; PubMed=11161025;  
RA Kingli C., Hauf G., Keller B.;  
RT "Hydrophobic properties of the structural protein GRP1.8 in the cell wall of protoxylem elements.";  
RL Plant Physiol. 125:673-682(2001).  
DR EMBL; AJ276509; CAB88804.1; -;  
DR InterPro; IPR002952; Eggshell.  
DR InterPro; IPR000817; Prion.  
DR PRINTS; PR01228; EGGSHLL.  
DR PRINTS; PR00341; PRION.  
KW Signal.  
FT SIGNAL. 1 28 POTENTIAL.  
FT CHAIN 29 390 GLYCINE-RICH PROTEIN GRP1.  
SQ SEQUENCE 390 AA; 28740 MW; 6EF5E2AA751A8B04 CRC64;

Query Match 2.2%; Score 8; DB 10; Length 390;  
Best Local Similarity 100.0%; Pred. No. 39; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
Db 83 AGGGAGGG 90

RESULT 44  
Q64217  
ID Q64217 PRELIMINARY; PRT; 395 AA.  
AC Q64217;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE Interphotoreceptor retinoid binding protein (Fragment).  
OS Spalax zemni (Podolsk mole rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;  
OC Spalax.  
OX NCBI\_TaxID=42150;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96304325; PubMed=8660440;  
RA Stanhope M.J., Smith M.R., Waddell V.G., Porter C.A., Shivji M.S.,  
RA Goodman M.;  
RT "Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades.";  
RL J. Mol. Evol. 43:83-92(1996).  
DR EMBL; U48589; AAB18667.1; -;  
DR InterPro; IPR003581; TSPC.  
DR Pfam; PF02692; IRBP; 1.

DR SMART; SM00245; TSPC; 1.  
KW Receptor.  
FT NON\_TER 1 1  
FT NON\_TER 395 395  
SQ SEQUENCE 395 AA; 42331 MW; F5391D0D4DD1D266 CRC64;

Query Match 2.2%; Score 8; DB 11; Length 395;  
Best Local Similarity 100.0%; Pred. No. 39; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 LEAPQQAP 109  
Db 53 LEAPQQAP 60

RESULT 45  
Q8GTL0  
ID Q8GTL0 PRELIMINARY; PRT; 400 AA.  
AC Q8GTL0;  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Putative glycine-rich cell wall protein precursor.  
GN QJ1165\_F02.18.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoidae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC clone:QJ1165\_F02.";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP003816; BAC21356.1; -;  
SQ SEQUENCE 400 AA; 28516 MW; 84B816F150070E60 CRC64;

Query Match 2.2%; Score 8; DB 10; Length 400;  
Best Local Similarity 100.0%; Pred. No. 40; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
Db 161 AGGGAGGG 168

RESULT 46  
Q9C518  
ID Q9C518 PRELIMINARY; PRT; 401 AA.  
AC Q9C518;  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE Auxin-induced basic helix-loop-helix transcription factor, putative.  
GN T18124.2 OR T15W6.11 OR AT1G58100.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=21016719; PubMed=11130712;  
RA Theologis A., Ecker J.R., Palm C.J., Pederspiel N.A., Kaul S.,  
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,  
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,  
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,

```

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaya I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militecher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Sallberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana";
RL Nature 408:816-820 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carrinci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC079131; AAG50759.1; -.
DR ENBL; AC079604; AAG50694.1; -.
DR EMBL; AY081344; AAL91233.1; -.
DR InterPro; IPR005333; TCP.
DR Pfam; PF03634; TCP; 1.
SQ SEQUENCE 401 AA; 42471 MW; 9F58817CDB11874C CRC64;

Query Match      2.2%; Score 8; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241
DB 294 AGGGAGGG 301

RESULT 47
Q8H010
ID Q8H010 PRELIMINARY; PRT; 420 AA.
AC Q8H010;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE Hypothetical protein.
GN Cl081D05:6.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]_TaxID=39947;
RP SEQUENCE FROM N.A.
RA Ming R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Currie J.,
RA Collura K.;
RT "Rice Genomic Sequence.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC134229; AAC06972.1; -.
DR Hypothetical protein.
SQ SEQUENCE 420 AA; 44730 MW; 344B8622E20ABFCE CRC64;

Query Match      2.2%; Score 8; DB 10; Length 420;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241
DB 81 AGGGAGGG 88

RESULT 48
Q8K1Q2
ID Q8K1Q2 PRELIMINARY; PRT; 423 AA.
AC Q8K1Q2;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Interphotoreceptor retinoid binding protein (Fragment).
GN IRBP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=wild type;
RA Huchon D., Madsen O., Sibbald M.J.J.B., Ament K., Stanhope M.J.,
RA Catzeflis F., de Jong W.W., Douzery E.J.P.;
RT "Rodent phylogeny and a timescale for the evolution of Glires:
RT evidence from an extensive taxon sampling using three nuclear genes.";
RL Mol. Biol. Evol. 0:0-0 (2002).
DR EMBL; AJ429134; CAD22102.1; -.
DR InterPro; IPR003581; TSPC.
DR Pfam; PF02692; IRBP; 2.
DR SMART; SM00245; TSPC; 1.
DR Receptor.
KW Receptor.
FT NON_TER 1
FT NON_TER 423
FT NON_TER 424
SQ SEQUENCE 423 AA; 46154 MW; DB96D43C18CA73E5 CRC64;

Query Match      2.2%; Score 8; DB 11; Length 423;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LEAQQOAP 109
DB 54 LEAQQOAP 61

RESULT 49
Q8K1P1
ID Q8K1P1 PRELIMINARY; PRT; 424 AA.
AC Q8K1P1;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Interphotoreceptor retinoid binding protein (Fragment).
GN IRBP.
OS Tachyoryctes sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Tachyoryctes.
OX NCBI_TaxID=123371;
RN [1]_TaxID=123371;
RP SEQUENCE FROM N.A.
RC STRAIN=#4991;
RA Huchon D., Madsen O., Sibbald M.J.J.B., Ament K., Stanhope M.,
RA Catzeflis F., de Jong W.W., Douzery E.J.P.;
RT "Rodent phylogeny and a timescale for the evolution of Glires:
RT evidence from an extensive taxon sampling using three nuclear genes.";
RL Mol. Biol. Evol. 0:0-0 (2002).
DR EMBL; AJ427231; CAD20269.1; -.
DR InterPro; IPR003581; TSPC.
DR Pfam; PF02692; IRBP; 2.
DR SMART; SM00245; TSPC; 1.
DR Receptor.
KW Receptor.
FT NON_TER 1
FT NON_TER 424
FT NON_TER 424
SQ SEQUENCE 424 AA; 45978 MW; 21EF2B620EB439F2 CRC64;

Query Match      2.2%; Score 8; DB 11; Length 424;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LEAQQOAP 109
DB 54 LEAQQOAP 61

```

```

Db          55 LEAPQAP 62
RESULT 50
Q9GR85      PRELIMINARY;      PRT;      427 AA.
AC Q9GR85;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE HrtLC2 protein.
GN HrtLC2.
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyridae; Halocynthia.
OX NCBI_taxID=7729;
[1]
RN
SEQUENCE FROM N.A.
RA Takahashi H., Satoh N.;
RT "Trunk lateral cell-specific genes of ascidian Halocynthia roretzi.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB053355; BAB20902.1; -.
DR HSPF; PI9339; ISXL.
DR InterPro; IPR002343; Hud_Sx1_RNA.
DR Pfam; PF00076; xrm; 2.
DR PRINTS; PR00961; HUDSXLRNA.
DR SMART; SMO0360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS0030; RRM_RNP_1; 2.
SQ SEQUENCE 427 AA; 46483 MW; 28017858775EB032 CRC64;

Query Match          2.2%; Score 8; DB 5; Length 427;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          85 TSSSQHSN 92
Db          37 TSSSQHSN 44

RESULT 51
Q9NG13      PRELIMINARY;      PRT;      438 AA.
AC Q9NG13;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Intermediate filament protein C2 (Fragment).
GN C2.
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_taxID=7739;
[1]
RN
SEQUENCE FROM N.A.
RX MEDLINE=20174564; PubMed=10711422;
RA Karabinos A., Riemer D., Panopoulou G., Lehrach H., Weber K.;
RT "Characterisation and tissue-specific expression of the two keratin
RT subfamilies of intermediate filament proteins in the cephalochordate
RT Branchiostoma.";
RL Eur. J. Cell Biol. 79:17-26(2000).
DR EMBL; AJ245429; CAB75938.1; -.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
FT NON_TER 1
FT NON_TER 438
SQ SEQUENCE 438 AA; 47162 MW; 5958FE10F6ED979B CRC64;

Query Match          2.2%; Score 8; DB 5; Length 438;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          234 AGGAGGG 241
Db          429 AGGAGGG 436

RESULT 52
Q8PJY7      PRELIMINARY;      PRT;      458 AA.
AC Q8PJY7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ATP-dependent RNA helicase.
GN DBPA OR XAC2390.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_taxID=92829;
[1]
RN
SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergio F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., Oliveira V.R.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.P.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AE011876; AM37242.1; -.
DR InterPro; IPR005580; DbpA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000629; DEAD_box.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF03880; DbpA; 1.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
KW Helicase; Complete proteome.
SQ SEQUENCE 458 AA; 49269 MW; 44A78F79B9BB06EC CRC64;

Query Match          2.2%; Score 8; DB 16; Length 458;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          210 TGEAGLSG 217
Db          407 TGEAGLSG 414

RESULT 53
Q8P8G0      PRELIMINARY;      PRT;      458 AA.
AC Q8P8G0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ATP-dependent RNA helicase.
GN DBPA OR XCC2282.

```

OS Xanthomonas campestris (pv. campestris).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=340;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33913 / NCPPB 528;  
 RX MEDLINE=22022145; PubMed=112024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Camnava F., Cardozo J., Chamberg P., Clapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities."  
 RL Nature 417:459-463(2002).  
 DR EMBL; AB012335; AAM41561.1; --  
 DR InterPro; IPR005580; DbpA.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR000629; DEAD\_box.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR Pfam; PF03880; DbpA; 1.  
 DR Pfam; PF00270; DEAD; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR PROSITE; PS00039; DEAD ATP HELICASE; 1.  
 KW Helicase; Complete proteome.  
 SQ SEQUENCE 458 AA; 49312 MW; CEE5277ACB98C438 CRC64;  
 Query Match 2.2%; Score 8; DB 16; Length 458;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 210 TGEAGLSG 217  
 Db 407 TGEAGLSG 414  
 RESULT 54  
 Q9UNW9 PRELIMINARY; PRT; 492 AA.  
 AC Q9UNW9; Q9UEAL;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE RNA-binding protein Nova-2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99007301; PubMed=9789075;  
 RX Yang Y.Y., Yin G.L., Darnell R.B.;  
 RT "The neuronal RNA-binding protein Nova-2 is implicated as the  
 RT autoantigen targeted in FOMIA patients with dementia."  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:13254-13259(1998).  
 RN [2]  
 RP SEQUENCE OF 29-492 FROM N.A.  
 RA Lamerdin J.E., McCredy P.M., Skowronski E., Viswanathan V.,  
 RA Burkhardt-Schultz K.J., Gordon L., Dias J., Ramirez M., Stilwagen S.,  
 RA Phan H., Velasco N., Do L., Regala N., Terry A., Barnes J.,  
 RA Danganan L., Brler A., Christensen M., Georgescu A., Avila J., Liu S.,  
 RA Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,

RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,  
 RA Arellano A., Brower A., Sanders C., Ow D., Nolan M., Trong S.,  
 RA Kobayashi A., Olsen A.S., Carriano A.V.;  
 RT "Sequence analysis of a 1.9 Mb region in 19q13.2 between APOE and  
 RT D19S412."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -|- SIMILARITY: CONTAINS 3 KH DOMAINS.  
 DR EMBL; AF083898; AAC72355.1; --  
 DR EMBL; AC006540; AAD13116.1; --  
 DR InterPro; IPR004087; KH dom.  
 DR InterPro; IPR004088; KH\_type\_1.  
 DR Pfam; PF00013; KH; 3.  
 DR SMART; SM00322; KH; 3.  
 DR PROSITE; PS00084; KH TYPE 1; 3.  
 SQ SEQUENCE 492 AA; 49008 MW; 41B63EAF6899256B CRC64;  
 Query Match 2.2%; Score 8; DB 4; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 234 AGGGAGGG 241  
 Db 370 AGGGAGGG 377  
 RESULT 55  
 Q35392 PRELIMINARY; PRT; 492 AA.  
 AC Q35392;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Forkhead 2.  
 GN FOXD2 OR MF2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Mesenchyme;  
 RX MEDLINE=98168839; PubMed=9510020;  
 RA Wu S.C.-Y., Grindley J., Winnier G.E., Hargrett L., Hogan B.L.M.;  
 RT "Mouse Mesenchyme forkhead 2 (Mf2): expression, DNA binding and  
 RT induction by sonic hedgehog during somitogenesis."  
 RL Mech. Dev. 70:3-13(1998).  
 DR EMBL; AF023915; AAB81275.1; --  
 DR HSP; Q63245; 2HPH.  
 DR TRANSFAC; T02492; --  
 DR MGD; MGI:1347471; Foxd2.  
 DR InterPro; IPR001766; TF Fork\_head.  
 DR Pfam; PF00250; Fork\_head; 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR ProDom; PD000425; TF Fork\_head; 1.  
 DR SMART; SM00339; FH; 1.  
 DR PROSITE; PS00657; FORK HEAD 1; 1.  
 DR PROSITE; PS00658; FORK HEAD 2; 1.  
 DR PROSITE; PS00039; FORK HEAD 3; 1.  
 SQ SEQUENCE 492 AA; 48936 MW; 7F82440F4C435702 CRC64;  
 Query Match 2.2%; Score 8; DB 11; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 234 AGGGAGGG 241  
 Db 392 AGGGAGGG 399  
 RESULT 56  
 Q43267 PRELIMINARY; PRT; 498 AA.  
 ID Q43267  
 AC Q43267;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE ASTROCYTIC NOVA1-like RNA-binding protein (NEUROONCOLOGIC ventral  
DE antigen 3) (Fragment).  
GN ANOVA OR NOVA3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20197319; PubMed=10735272;  
RA Ueki K., Ramaswamy S., Billings S.J., Mohrenweiser H.W., Louis D.N.;  
RT "ANOVA, a putative astrocytic RNA binding protein gene that maps to  
RT chromosome 19q13.3." (1997).  
RL Neurogenetics 1:31-36(1997).  
CC -!- FUNCTION: MAY REGULATE RNA SPLICING OR METABOLISM IN A SPECIFIC  
CC SUBSET OF DEVELOPING NEURONS (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: BRAIN. EXPRESSION RESTRICTED TO ASTROCYTES.  
CC -!- DISEASE: MAY BE A TARGET ANTIGEN IN ONE OF THE UNDEFINED HUMAN  
CC PARANEOPLASTIC SYNDROMES.  
DR EMBL; U70477; AAB8661.1; -.  
DR Genew; HGNC:7887; NOVA2.  
DR InterPro; IPR004087; KH dom.  
DR InterPro; IPR004088; KH\_type\_1.  
DR Pfam; PF00013; KH; 3.  
DR SMART; SM00322; KH; 3.  
DR PROSITE; PS00084; KH TYPE 1; 3.  
KW Nuclear protein; RNA-binding; Repeat; Antigen.  
FT NON\_TER 1  
FT DOMAIN 1 76 KH.  
FT DOMAIN 140 174 KH.  
FT DOMAIN 246 255 POLY-ALA.  
FT DOMAIN 325 330 POLY-ALA.  
FT DOMAIN 350 356 POLY-PRO.  
FT DOMAIN 363 368 POLY-ALA.  
FT DOMAIN 375 385 POLY-GLY.  
FT DOMAIN 389 397 POLY-ALA.  
FT DOMAIN 416 446 KH.  
SQ SEQUENCE 498 AA; 49721 MW; C4B54196FDB6BF78 CRC64;

Query Match 2.2%; Score 8; DB 4; Length 498;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
Db 376 AGGGAGGG 363  
|||||

RESULT 57  
Q816C7  
ID Q816C7 PRELIMINARY; PRT; 544 AA.  
AC Q816C7;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CG3588 (Fragment).  
GN CG3588.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Zimbabwe/Harare19;  
RX MEDLINE=22247773; PubMed=12351680;  
RA Harr B., Kauer M., Schlotterer C.;  
RT "Hitchhiking mapping: A population-based fine-mapping strategy for

RT adaptive mutations in Drosophila melanogaster.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:12949-12954 (2002).  
DR EMBL; AF495952; AAN09997.1; -.  
FT NON\_TER 1  
FT NON\_TER 544 544  
SQ SEQUENCE 544 AA; 55701 MW; 640FA3F264821E75 CRC64;  
Query Match 2.2%; Score 8; DB 5; Length 544;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 NAGGGAGG 240  
Db 516 NAGGGAGG 523  
|||||

RESULT 58  
Q816C6  
ID Q816C6 PRELIMINARY; PRT; 544 AA.  
AC Q816C6;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CG3588 (Fragment).  
GN CG3588.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Zimbabwe/Harare20;  
RX MEDLINE=22247773; PubMed=12351680;  
RA Harr B., Kauer M., Schlotterer C.;  
RT "Hitchhiking mapping: A population-based fine-mapping strategy for  
RT adaptive mutations in Drosophila melanogaster.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:12949-12954 (2002).  
DR EMBL; AF495953; AAN09998.1; -.  
FT NON\_TER 1  
FT NON\_TER 544 544  
SQ SEQUENCE 544 AA; 55771 MW; 9FDA1909E06A17AB CRC64;

Query Match 2.2%; Score 8; DB 5; Length 544;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 NAGGGAGG 240  
Db 516 NAGGGAGG 523  
|||||

RESULT 59  
Q816C5  
ID Q816C5 PRELIMINARY; PRT; 544 AA.  
AC Q816C5;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CG3588 (Fragment).  
GN CG3588.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kenia11;  
RX MEDLINE=22247773; PubMed=12351680;  
RA Harr B., Kauer M., Schlotterer C.;  
RT "Hitchhiking mapping: A population-based fine-mapping strategy for  
RT adaptive mutations in Drosophila melanogaster.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:12949-12954 (2002).  
 DR EMBL; AF495954; AAN09999.1; -.  
 FT NON TER 1 1  
 FT NON TER 544 544  
 SQ SEQUENCE 544 AA; 55757 MW; 774BD4F723C22386 CRC64;

Query Match 2.2%; Score 8; DB 5; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NAGGGAGG 240  
 |||||  
 Db 516 NAGGGAGG 523

## RESULT 60

ID Q816C4 PRELIMINARY; PRT; 544 AA.  
 AC Q816C4;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE CG3588 (Fragment).  
 GN CG3588.

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Kenial14;  
 RX MEDLINE=22247773; PubMed=12351680;

RA Harr B., Kauer M., Schloetterer C.;  
 RT "Hitchhiking mapping: A population-based fine-mapping strategy for  
 RT adaptive mutations in Drosophila melanogaster.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12949-12954 (2002).  
 DR EMBL; AF495955; AAN10000.1; -.  
 FT NON TER 1 1  
 FT NON TER 544 544  
 SQ SEQUENCE 544 AA; 55691 MW; 619CB335141D9389 CRC64;

Query Match 2.2%; Score 8; DB 5; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NAGGGAGG 240  
 |||||  
 Db 516 NAGGGAGG 523

## RESULT 61

ID Q816C3 PRELIMINARY; PRT; 544 AA.  
 AC Q816C3;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE CG3588 (Fragment).  
 GN CG3588.

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Kenial17;  
 RX MEDLINE=22247773; PubMed=12351680;

RA Harr B., Kauer M., Schloetterer C.;  
 RT "Hitchhiking mapping: A population-based fine-mapping strategy for  
 RT adaptive mutations in Drosophila melanogaster.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12949-12954 (2002).

DR EMBL; AF495956; AAN10001.1; -.  
 FT NON TER 1 1  
 FT NON TER 544 544  
 SQ SEQUENCE 544 AA; 55757 MW; 4ADF86A1215BEEA3 CRC64;

Query Match 2.2%; Score 8; DB 5; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NAGGGAGG 240  
 |||||  
 Db 516 NAGGGAGG 523

## RESULT 62

ID Q816B7 PRELIMINARY; PRT; 544 AA.  
 AC Q816B7;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE CG3588 (Fragment).  
 GN CG3588.

OS Drosophila simulans (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7240;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=New Caledonia;  
 RX MEDLINE=22247773; PubMed=12351680;

RA Harr B., Kauer M., Schloetterer C.;  
 RT "Hitchhiking mapping: A population-based fine-mapping strategy for  
 RT adaptive mutations in Drosophila melanogaster.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12949-12954 (2002).  
 DR EMBL; AF495972; AAN10017.1; -.  
 FT NON TER 1 1  
 FT NON TER 544 544  
 SQ SEQUENCE 544 AA; 55449 MW; 9CF8E3EB9EB7A755 CRC64;

Query Match 2.2%; Score 8; DB 5; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NAGGGAGG 240  
 |||||  
 Db 516 NAGGGAGG 523

## RESULT 63

ID Q81072 PRELIMINARY; PRT; 544 AA.  
 AC Q81072;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE CG3588 (Fragment).  
 GN CG3588.

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Zimbabwe/Harare12, and Zimbabwe/Harare18;  
 RX MEDLINE=22247773; PubMed=12351680;

RA Harr B., Kauer M., Schloetterer C.;  
 RT "Hitchhiking mapping: A population-based fine-mapping strategy for  
 RT adaptive mutations in Drosophila melanogaster.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12949-12954 (2002).  
 DR EMBL; AF495949; AAN09994.1; -.

DR EMBL: AF495951; AAN09996.1; --.  
 FT NON\_TER 1 1  
 FT NON\_TER 544 544  
 SQ SEQUENCE 544 AA; 55729 MW; DEDALCODE16ALED2 CRC64;

Query Match 2.2%; Score 8; DB 5; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 NAGGAGG 240  
 Db 516 NAGGAGG 523  
 |||||

RESULT 64

Q22809 PRELIMINARY; PRT; 548 AA.  
 AC Q22809;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical 61.8 kDa protein.  
 GN T26C11.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderiinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Martin J.;  
 RT "The sequence of C. elegans cosmid T26C11.";  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U41017; AAC48213.1; --.  
 DR WormPep; T26C11.4; CE05023.  
 KW Hypothetical protein.  
 SQ SEQUENCE 548 AA; 61841 MW; 0BF798A10B148E03 CRC64;

Query Match 2.2%; Score 8; DB 5; Length 548;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 PQGASNAG 235  
 Db 195 PQGASNAG 202  
 |||||

RESULT 65

Q81GB4 PRELIMINARY; PRT; 550 AA.  
 ID Q81GB4;  
 AC Q81GB4;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE RH70633p.  
 GN CG11711.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=y;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BT001864; AAN71631.1; --.  
 SQ SEQUENCE 550 AA; 57774 MW; 36A7B8CAD285E412 CRC64;

Query Match 2.2%; Score 8; DB 5; Length 550;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGG 241  
 Db 95 AGGGAGG 102  
 |||||

RESULT 66

Q9VTJ7 PRELIMINARY; PRT; 566 AA.  
 ID Q9VTJ7;  
 AC Q9VTJ7;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE CG11711 protein.  
 GN CG11711 OR CG11712.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., M.G.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,



RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195(2000).  
DR EMBL; AF003544; AAF50051.1; -.  
DR FlyBase; FBgn0036172; CG11711.  
DR InterPro; IPR005301; Mobl\_phocoin.  
DR Pfam; PF03637; Mobl\_phocoin; 1.  
SQ SEQUENCE 566 AA; 58907 MW; 1D3F57A2659DEC2C CRC64;

Query Match 2.2%; Score 8; DB 5; Length 566;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
Db 95 AGGGAGGG 102  
|||||

RESULT 67  
Q8LRM4 PRELIMINARY; PRT; 579 AA.  
AC Q8LRM4;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative amt protein.  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-9;  
RA Yao T., Feild E., King N., Soupen E., Kim K.-S., Kustu S.;  
RT "Chlamydomonas reinhardtii amt gene."  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF509497; AAM43911.1; -.  
DR InterPro; IPR001905; Ammonium\_transpt.  
DR Pfam; PF00909; Ammonium\_transp; 1.  
DR TIGRFAMs; TIGR00836; amt; 1.  
DR PROSITE; PS01219; AMMONIUM\_TRANS; 1.  
SQ SEQUENCE 579 AA; 60697 MW; 170C49CF09FB5C2D CRC64;

Query Match 2.2%; Score 8; DB 10; Length 579;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
Db 535 AGGGAGGG 542  
|||||

RESULT 68  
Q8LRM5 PRELIMINARY; PRT; 579 AA.  
AC Q8LRM5;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative amt protein.  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CC-125;  
RA Yao T., Feild E., King N., Soupen E., Kim K.-S., Kustu S.;  
RT "Chlamydomonas reinhardtii amt gene.";

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF509496; AAM43910.1; -.  
DR InterPro; IPR001905; Ammonium\_transpt.  
DR Pfam; PF00909; Ammonium\_transp; 1.  
DR TIGRFAMs; TIGR00836; amt; 1.  
DR PROSITE; PS01219; AMMONIUM\_TRANS; 1.  
SQ SEQUENCE 579 AA; 60683 MW; B20C4CCA09FB5C2B CRC64;

Query Match 2.2%; Score 8; DB 10; Length 579;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
Db 535 AGGGAGGG 542  
|||||

RESULT 69  
Q9FCJ3 PRELIMINARY; PRT; 584 AA.  
AC Q9FCJ3;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein SC05195.  
GN SC05195 OR 2SC3B6.19.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Cardeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb *Streptomyces coelicolor* A3(2) chromosome."  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete *Streptomyces  
coelicolor* A3(2)."  
RL Nature 417:141-147(2002).  
DR EMBL; AL939122; CAC01325.1; -.  
DR InterPro; IPR002838; DUF124.  
DR InterPro; IPR002965; P\_rich\_extensn.  
DR InterPro; IPR003325; TerD.  
DR Pfam; PF02342; TerD; 1.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR ProDom; PD013634; DUF124; 1.  
KW Hypothetical protein; Complete proteome.



AC 062006;  
DT 01-AUG-1998 (TRENBLrel. 07, Created)  
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Intermediate filament protein C2.  
OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).  
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
OC Branchiostoma.  
OX NCBI\_TaxID=7740;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98267218; PubMed=9602172;  
RA Riemer D., Karabinos A., Weber K.;  
RT "Analysis of eight cDNAs and six genes for intermediate filament (IF) proteins in the cephalochordate Branchiostoma reveals differences in the IF multigene families of lower chordates and the vertebrates.";  
RL Gene 211:361-373(1998).  
DR EMBL; AJ223578; CAA11445.1; -;  
DR InterPro; IPR002952; Eggshell.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_1.  
DR Pfam; PF00038; filament; 5.  
DR PRINTS; PRO1228; EGGSHLL.  
DR PROSITE; PRO1248; TYPE1KERATIN.  
DR PROSITE; PS00226; IF; 1.  
SQ SEQUENCE 795 AA; 84532 MW; D00B2665CEACTFFC CRC64;  
Query Match 2.2%; Score 8; DB 5; Length 795;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 234 AGGAGGG 241  
Db 492 AGGAGGG 499  
|||||  
RESULT 74  
Q43715  
ID Q43715 PRELIMINARY; PRT; 809 AA.  
AC Q43715;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Chloroplastic outer envelope membrane protein (OEP75) precursor (OEP75).  
DE (OEP75).  
GN IAP75.  
OS Pisum sativum (Garden pea).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids 1; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.  
OX NCBI\_TaxID=3888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=CV. LITTLE MARVEL; TISSUE=Leaf;  
RX MEDLINE=95300776; PubMed=7781598;  
RA Tranel P.J., Froehlich J., Goyal A., Keegstra K.;  
RT "A component of the chloroplastic protein import apparatus is targeted to the outer envelope membrane via a novel pathway.";  
RL EMBO J. 14:2436-2446(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95063931; PubMed=7973649;  
RA Schnell D.J., Kessler P., Blobel G.;  
RT "Isolation of components of the chloroplast protein import machinery.";  
RL Science 266:1007-1012(1994).  
DR EMBL; X83767; CAA58720.1; -;  
DR EMBL; L36858; AAA53275.1; -;  
DR InterPro; IPR000184; Bac surfAg\_D15.  
DR InterPro; IPR005689; IAP75.  
DR Pfam; PF01103; Bac\_surface\_Ag; 1.  
DR TIGRFAWS; TIGR00992; 3a0901S03IAP75; 1.  
KW Chloroplast; Outer membrane; Transit peptide.

FT TRANSIT 1 131 POTENTIAL.  
FT CHAIN 132 809 CHLOROPLASTIC OUTER ENVELOPE MEMBRANE  
FT PROTEIN (OEP75).  
SQ SEQUENCE 809 AA; 88269 MW; AFES1AE75F0617C5 CRC64;  
Query Match 2.2%; Score 8; DB 10; Length 809;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 234 AGGAGGG 241  
Db 90 AGGAGGG 97  
|||||  
RESULT 75  
Q25343  
ID Q25343 PRELIMINARY; PRT; 823 AA.  
AC Q25343;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE L2759.4.  
GN L2759.4.  
OS Leishmania major.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Friedlin;  
RX MEDLINE=99178987; PubMed=10077609;  
RA Myler P.J., Audleman L., deVos T., Hixson G., Kiser P., Lenley C.,  
RA Magness C., Rickel E., Sisk E., Sunkin S., Swartzell S., Westlake T.,  
RA Bastien P., Fu G., Ivens A., Stuart K.;  
RT "Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-coding genes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:2902-2906(1999).  
DR EMBL; AE001274; AAC24635.1; -;  
DR InterPro; IPR006626; PbH1.  
DR SMART; SM00710; PbH1; 4.  
SQ SEQUENCE 823 AA; 86775 MW; 5631E7DFC1A70435 CRC64;  
Query Match 2.2%; Score 8; DB 5; Length 823;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 56 ESTLEENK 63  
Db 736 ESTLEENK 743  
|||||  
Search completed: December 17, 2003, 06:35:19  
Job time : 41.1811 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:20:56 ; Search time 34.7529 Seconds  
(without alignments)  
1680.762 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 368

Sequence: 1 MKLTVKTLGTHPFRVQPN.....CDRNEELAAVYLLHAGEED 368

Scoring table:

Gapop 60.0 ; Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : A.Geneseq 19Jun03.\*

1: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*  
24: /SIDSI1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	368	100.0	368	21	Maize Rad23 protei
2	56	15.2	117	23	Human ORF3591 prot
3	21	5.7	299	21	Arabidopsis thalia
4	21	5.7	348	21	Arabidopsis thalia
5	21	5.7	371	21	Arabidopsis thalia
6	15	4.1	186	21	Arabidopsis thalia
7	15	4.1	214	21	Arabidopsis thalia
8	15	4.1	257	21	Arabidopsis thalia
9	15	4.1	296	21	Arabidopsis thalia

10	15	4.1	307	21	AA19976	Arabidopsis thalia
11	15	4.1	307	21	AA45206	Arabidopsis thalia
12	15	4.1	335	21	AA19975	Arabidopsis thalia
13	15	4.1	335	21	AA45205	Arabidopsis thalia
14	15	4.1	345	21	AA44342	Arabidopsis thalia
15	15	4.1	368	21	AA44341	Arabidopsis thalia
16	15	4.1	378	21	AA19974	Arabidopsis thalia
17	15	4.1	378	21	AA45204	Arabidopsis thalia
18	13	3.5	405	21	AA171458	Maize Rad23 protei
19	11	3.0	136	21	AA19942	Arabidopsis thalia
20	11	3.0	414	22	AB58584	Arabidopsis thalia
21	9	2.4	92	22	AB16631	Novel human diagn
22	8	2.2	10	22	AA19974	Arabidopsis thalia
23	8	2.2	10	22	AA19974	Arabidopsis thalia
24	8	2.2	60	23	AA19974	Arabidopsis thalia
25	8	2.2	174	21	AA19974	Arabidopsis thalia
26	8	2.2	265	19	AA19974	Arabidopsis thalia
27	8	2.2	325	10	AA19974	Arabidopsis thalia
28	8	2.2	325	18	AA19974	Arabidopsis thalia
29	8	2.2	331	24	AA19974	Arabidopsis thalia
30	8	2.2	332	21	AA19974	Arabidopsis thalia
31	8	2.2	352	11	AA19974	Arabidopsis thalia
32	8	2.2	566	22	AB58587	Arabidopsis thalia
33	8	2.2	604	21	AA19974	Arabidopsis thalia
34	8	2.2	620	22	AB58588	Arabidopsis thalia
35	8	2.2	712	11	AA19974	Arabidopsis thalia
36	8	2.2	738	19	AA19974	Arabidopsis thalia
37	8	2.2	828	23	AB58589	Arabidopsis thalia
38	8	2.2	828	24	AB58589	Arabidopsis thalia
39	8	2.2	1004	22	AA19974	Arabidopsis thalia
40	8	2.2	1088	24	AB58589	Arabidopsis thalia
41	8	2.2	1329	23	AA19974	Arabidopsis thalia
42	8	2.2	1385	22	AB58589	Arabidopsis thalia
43	8	2.2	1417	22	AB58589	Arabidopsis thalia
44	8	2.2	2075	22	AB58589	Arabidopsis thalia
45	7	1.9	9	22	AA19974	Arabidopsis thalia
46	7	1.9	9	22	AA19974	Arabidopsis thalia
47	7	1.9	10	22	AA19974	Arabidopsis thalia
48	7	1.9	10	22	AA19974	Arabidopsis thalia
49	7	1.9	10	22	AA19974	Arabidopsis thalia
50	7	1.9	10	22	AA19974	Arabidopsis thalia
51	7	1.9	10	22	AA19974	Arabidopsis thalia
52	7	1.9	10	22	AA19974	Arabidopsis thalia
53	7	1.9	10	22	AA19974	Arabidopsis thalia
54	7	1.9	10	22	AA19974	Arabidopsis thalia
55	7	1.9	12	17	AA19974	Arabidopsis thalia
56	7	1.9	14	13	AA19974	Arabidopsis thalia
57	7	1.9	14	13	AA19974	Arabidopsis thalia
58	7	1.9	15	7	AA19974	Arabidopsis thalia
59	7	1.9	15	24	AB58589	Arabidopsis thalia
60	7	1.9	15	24	AB58589	Arabidopsis thalia
61	7	1.9	15	24	AB58589	Arabidopsis thalia
62	7	1.9	15	24	AB58589	Arabidopsis thalia
63	7	1.9	15	24	AB58589	Arabidopsis thalia
64	7	1.9	15	24	AB58589	Arabidopsis thalia
65	7	1.9	15	24	AB58589	Arabidopsis thalia
66	7	1.9	16	13	AA19974	Arabidopsis thalia
67	7	1.9	16	13	AA19974	Arabidopsis thalia
68	7	1.9	17	13	AA19974	Arabidopsis thalia
69	7	1.9	17	13	AA19974	Arabidopsis thalia
70	7	1.9	17	13	AA19974	Arabidopsis thalia
71	7	1.9	19	23	AB58589	Arabidopsis thalia
72	7	1.9	20	7	AA19974	Arabidopsis thalia
73	7	1.9	20	11	AA19974	Arabidopsis thalia
74	7	1.9	20	13	AA19974	Arabidopsis thalia
75	7	1.9	20	13	AA19974	Arabidopsis thalia
76	7	1.9	20	13	AA19974	Arabidopsis thalia
77	7	1.9	25	11	AA19974	Arabidopsis thalia
78	7	1.9	25	11	AA19974	Arabidopsis thalia
79	7	1.9	35	15	AA19974	Arabidopsis thalia
80	7	1.9	55	22	AB58589	Arabidopsis thalia
81	7	1.9	55	22	AB58589	Arabidopsis thalia
82	7	1.9	55	22	AB58589	Arabidopsis thalia

83 7 1.9 55 22 AB23322 Protein #5321 enco  
 84 7 1.9 55 22 AAM58743 Human brain expres  
 85 7 1.9 55 22 AAM71250 Human bone marrow  
 86 7 1.9 55 22 AAM18952 Peptide #5186 enco  
 87 7 1.9 55 22 AAM31531 Peptide #5568 enco  
 88 7 1.9 55 23 ABG41049 Human peptide enco  
 89 7 1.9 56 23 ABP29010 Streptococcus poly  
 90 7 1.9 57 17 AAR88669 Human cytomagalovi  
 91 7 1.9 61 19 AAW75013 Human secreted pro  
 92 7 1.9 61 23 ABG95475 Human novel secret  
 93 7 1.9 61 23 ABP05846 Human ORFX protein  
 94 7 1.9 70 21 AAG09626 Arabidopsis thalia  
 95 7 1.9 72 23 ABP03541 Human ORFX protein  
 96 7 1.9 78 23 ABG60309 Lymphona associate  
 97 7 1.9 78 23 ABG60310 Lymphona associate  
 98 7 1.9 80 21 AAG57031 Arabidopsis thalia  
 99 7 1.9 81 20 AAV32776 HIV SP-2 strain ch  
 100 7 1.9 82 20 AAY32775 HIV SP-2 strain ch

## ALIGNMENTS

## RESULT 1

AA71459  
 ID AAY71459 standard; Protein; 368 AA.

XX AC AAY71459;  
 XX DT 04-OCT-2000 (first entry)  
 XX DE Maize Rad23 protein #2.

XX KW Rad23; maize; ATCC No: PTA-531; recombinant expression cassette; wheat;  
 XX KW transgenic plant; soybean; sunflower; canola; modulator.  
 XX OS Zea mays.

XX PN WO200031268-A1.  
 XX PD 02-JUN-2000.  
 XX PF 12-OCT-1999; 99WO-US24129.

XX PR 23-NOV-1998; 98US-0109728.  
 XX PA (PION-) PIONEER HI-BRED INT INC.

XX PI Mahajan PB, Tagliani L;  
 XX DR WPI; 2000-400078/34.  
 XX DR N-PSDB; AAD01231.

XX PT Isolated nucleic acid encoding maize RAD23 protein is used to modulate  
 XX PT the levels of polypeptides in plant or in assays for identifying  
 XX PT compounds that bind to and/or increase/decrease enzymatic activity of  
 XX PT catalytically active polypeptides -  
 XX PS Claim 11b; Page 78-79; 82pp; English.

XX CC The present sequence is the maize Rad23 protein #2. It is isolated from  
 XX CC a Zea mays cell line, B73 callus tissue regenerated five days after  
 XX CC transfer of the callus from medium containing auxin to a medium devoid  
 XX CC of exogenous auxin. The cDNA is deposited under the ATCC No: PTA-531.  
 XX CC Maize Rad23 DNA sequence operably linked to a promoter, can be used to  
 XX CC construct a recombinant expression cassette. This expression cassette  
 XX CC can be used to generate a dicot or monocot transgenic plant e.g., maize,  
 XX CC soybean, sunflower, sorghum, canola, wheat, etc. It can also be used to  
 XX CC modulate the levels of Rad23 polypeptide expression in a plant or in  
 XX CC assays to identify compounds, that bind to and/or modulate the enzymatic  
 XX CC activity of catalytically active polypeptides.

XX SQ Sequence 368 AA;

Query Match 100.0%; Score 368; DB 21; Length 368;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLTVKTLKGTHTFEIRVQPNNDTMAVKKNIIEIQKDSYPMQQLLIINGKVKLDESTLE 60  
 DB 1 MKLTVKTLKGTHTFEIRVQPNNDTMAVKKNIIEIQKDSYPMQQLLIINGKVKLDESTLE 60  
 QY 61 ENKVNEDGLVVMLSKGTSGTSSSHSNTPATROAPPLEAPQAPQPPVAPITTSQ 120  
 DB 61 ENKVNEDGLVVMLSKGTSGTSSSHSNTPATROAPPLEAPQAPQPPVAPITTSQ 120  
 QY 121 PEGLPAPQPNTHDAAASNLLSGRNVDTIINQLEMGGSGWDKQVQRAIRAAYNPNPERAV 180  
 DB 121 PEGLPAPQPNTHDAAASNLLSGRNVDTIINQLEMGGSGWDKQVQRAIRAAYNPNPERAV 180  
 QY 181 EYLYSGIPVTABIAVPIGGQANTTDRAPTGAAGSGIPTAPLDLPFGASNAGGAGG 240  
 DB 181 EYLYSGIPVTABIAVPIGGQANTTDRAPTGAAGSGIPTAPLDLPFGASNAGGAGG 240  
 QY 241 GPDLELRNNPQFQAVREVMHTNPQILQPMVLSEKONPQILRLIEENHDEFLLNEPPE 300  
 DB 241 GPDLELRNNPQFQAVREVMHTNPQILQPMVLSEKONPQILRLIEENHDEFLLNEPPE 300  
 QY 301 GGEGDFLOQPEDEMPHAI SVTPPEQEAIGRLSEMGFDRARVIEAPLACDRNEELAANYL 360  
 DB 301 GGEGDFLOQPEDEMPHAI SVTPPEQEAIGRLSEMGFDRARVIEAPLACDRNEELAANYL 360  
 QY 361 LEHAGEED 368  
 DB 361 LEHAGEED 368

## RESULT 2

ABP34618  
 ID ABP34618 standard; Protein; 117 AA.

XX AC ABP34618;  
 XX DT 08-JUL-2002 (first entry)  
 XX DE Human ORF3591 protein, SEQ ID NO: 7182.

XX KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; activin; inhibitor; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neotropic;  
 KW neuroprotective; antilatherosclerotic; anticoagulant; thrombolytic;  
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.

OS Homo sapiens.

XX WO200190366-A2.

XX PD 29-NOV-2001.

XX PF 24-MAY-2001; 2001WO-US17076.

XX PR 24-MAY-2000; 2000US-206690P.

XX PA (CURA-) CURAGEN CORP.

XX PI Leach MD, Shimkets RA;

XX DR WPI; 2002-106200/14.

DR	N-PSDS; ABN78644.
XX	
PT	Novel human polypeptides and polynucleotides useful for diagnosing,
PT	preventing and treating cardiovascular disease, neurodegenerative,
PT	hyperproliferative disorders and disorders related to organ
PT	transplantation -
XX	
XX	Claim 10; Page 2042; 2508pp; English.
XX	
CC	Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC	designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC	ABN79587 represent cDNAs encoding them. The invention also encompasses
CC	polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC	referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC	the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC	polynucleotides, the recombinant production of ORFX proteins, antibodies
CC	specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC	polypeptides, methods of screening for modulators of ORFX expression or
CC	activity, and methods of screening individuals for a predisposition to an
CC	ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC	range of biological activities, such as cytokine, cell proliferation,
CC	cell differentiation, immune modulation, haematopoiesis regulation,
CC	tissue growth, angiogenesis, activin or inhibin activity, chemotactic/ chemokinetic activity, haemostatic activity, thrombolytic activity, receptor/ligand, anti-inflammatory activity, tumour inhibition activity, and antitumour activity, and may also be involved in the determination of bodily characteristics, fertility and behaviour. ORFX proteins, nucleic acids and antibodies may be used in the treatment of cancers, other proliferative disorders such as psoriasis and benign tumours, neurological disorders such as epilepsy and Alzheimer's disease, cardiovascular diseases, immune system disorders, disorders related to organ transplantation, disorders of tissue growth and regeneration, diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester storage disease, and infectious diseases caused by viral, bacterial, fungal and other pathogens. ORFX nucleic acids may also be used as a source of primers and probes, in the detection of ORFX genomic sequences or transcripts, in the identification and cloning of homologous sequences, in genetic diagnosis, and in forensic biology. The ORFX nucleic acids may additionally be used to produce transgenic animals which may be useful for studying the function and/or activity of ORFX protein, and in drug screening. The ORFX proteins may also be used as immunogens to generate specific antibodies, which are useful in the diagnosis, treatment and monitoring of ORFX-associated diseases.
XX	
XX	
SQ	Sequence 117 AA;
	Query Match 15.2%; Score 56; DB 23; Length 117;
	Best Local Similarity 100.0%; Pred.No. 2.5e-47;
	Matches 56; Conservative 0; Mismatchles 0; Indels 0; Gaps 0
Qy	253 QAVREVHTNPQIIQLPMVLVELSKONPQILRLTEENHDEFLQLLNPEFGSGDFFLD 308
Dd	 29 QAVREVHTNPQIIQLPMVLVELSKONPQILRLTEENHDEFLQLLNPEFGSGDFFLD 84
RESULT 3	
ID AAG17379	
AD AAG17379 standard; Protein; 299 AA.	
XX	
AC AAG17379;	
XX	
DT 17-OCT-2000 (first entry)	
XX	
DE Arabidopsis thaliana protein fragment SEQ ID NO: 18374.	
XX	
KW Protein identification; signal transduction pathway; metabolic pathway;	
KW hybridisation assay; genetic mapping; gene expression control; promoter;	
XX termination sequence.	
OS Arabidopsis thaliana.	
XX	
PN EP1033405-A2.	
XX	

PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157553.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 5.7%; Score 21; DB 21; Length 299;

Best Local Similarity 100.0%; Pred. No. 4e-12;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 342 VIEAFACDRNEELAANYLLE 362  
Db 272 VIEAFACDRNEELAANYLLE 292

RESULT 4

AAG17378  
ID AAG17378 standard; Protein; 348 AA.

XX  
AC AAG17378;

XX  
DT 17-OCT-2000 (first entry)

XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 18373.

XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX  
OS Arabidopsis thaliana.

XX  
PN EP1033405-A2.

XX  
PD 06-SEP-2000.

XX  
PF 25-FEB-2000; 2000EP-0301439.

XX  
PR 25-FEB-1999; 99US-0121825.

PR  
05-MAR-1999; 99US-0123180.

PR  
09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149772.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150586.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151343.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.



```
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match 5.7% Score 21; DB 21; Length 348;
Best Local Similarity 100.0%; Pred. No. 4.6e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 VIEAFLACDRNEELANYLLE 362
   |||||
Db 321 VIEAFLACDRNEELANYLLE 341

RESULT 5
AAG17377
ID AAG17377 standard; Protein; 371 AA.
XX
AC AAG17377;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 18372.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
```

PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157531.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 5.7%; Score 21; DB 21; Length 371;  
Best Local Similarity 100.0%; Pred. No. 4.9e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 342 VIEAFLACDRNEELAAANYLLE 362  
|||||  
Db 344 VIEAFLACDRNEELAAANYLLE 364

## RESULT 6

AAG36531  
ID AAG36531 standard; Protein; 186 AA.

XX AC AAG36531;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 44781.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

PP 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.



PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 4.1%; Score 15; DB 21; Length 186;

Best Local Similarity 100.0%; Pred. No. 2.4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 NNPERAVEYLYSGIP 188  
 |||||  
 DB 106 NNPERAVEYLYSGIP 120

## RESULT 7

AAG36530  
 ID AAG36530 standard; Protein; 214 AA.

XX AC AAG36530;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 44780.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX FN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0128845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144894.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149168.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 16-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155859.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157173.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 4.1%; Score 15; DB 21; Length 214;  
Best Local Similarity 100.0%; Pred. No. 2.7e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 174 NNPERAVEVLYSGIP 188  
Db 134 NNPERAVEVLYSGIP 148  
|||||||  
  
RESULT 8  
AAG36529  
ID AAG36529 standard; Protein; 257 AA.  
AC AAG36529;  
XX  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 44779.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 09-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145921.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-01511303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.

```
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      4.1%; Score 15; DB 21; Length 257;
Best Local Similarity 100.0%; Pred. NO. 3.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 NNPERAVEVLYSGIP 188
DB 177 NNPERAVEVLYSGIP 191
|||||
|||||

RESULT 9
AAG44343
ID AAG44343 standard; Protein; 296 AA.
XX
AC AAG44343;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55533.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
```

PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155133.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 4.1%; Score 15; DB 21; Length 296;  
Best Local Similarity 100.0%; Pred. No. 3.7e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 349 CDRNEELAANYLLEH 363  
Db 276 CDRNEELAANYLLEH 290  
RESULT 10  
AAG19976  
ID AAG19976 standard; Protein; 307 AA.  
XX  
AC AAG19976;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21983.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 25-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.



PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144321.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145318.  
PR 27-JUL-1999; 99US-0145319.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150568.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162142.

Query Match 4.1%; Score 15; DB 21; Length 307;

Best Local Similarity 100.0%; Pred.No. 3.8e-06; Mismatches 0; Indels 0; Gaps 0;  
Matches 15; Conservative 0;

QY 174 NNPERAVEYLXSGIP 188  
Db 106 NNPERAVEYLXSGIP 120

RESULT 11

AA45206  
ID AA45206 standard; Protein; 307 AA.  
XX AC AA45206;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 56725.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridization assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145132.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 11-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149446.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

```
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.1%; Score 15; DB 21; Length 307;
Best Local Similarity 100.0%; Pred. No. 3.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 174 NNPERAVEVLYSGIP 188
    |||||
Db 106 NNPERAVEVLYSGIP 120

RESULT 12
AAG19975
ID AAG19975 standard; Protein; 335 AA.
XX
AC AAG19975;
XX
DT 17-OCT-2000 (first entry)
XX
```

DE Arabidopsis thaliana protein fragment SEQ ID NO: 21982.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899;  
PR 23-JUN-1999; 99US-0140353;  
PR 23-JUN-1999; 99US-0140354;  
PR 24-JUN-1999; 99US-0140695;  
PR 28-JUN-1999; 99US-0140823;  
PR 29-JUN-1999; 99US-0140991;  
PR 30-JUN-1999; 99US-0141287;  
PR 01-JUL-1999; 99US-0141842;  
PR 01-JUL-1999; 99US-0142154;  
PR 02-JUL-1999; 99US-0142055;  
PR 06-JUL-1999; 99US-0142390;  
PR 08-JUL-1999; 99US-0142803;  
PR 09-JUL-1999; 99US-0142920;  
PR 12-JUL-1999; 99US-0142977;  
PR 13-JUL-1999; 99US-0143542;  
PR 14-JUL-1999; 99US-0143624;  
PR 15-JUL-1999; 99US-0144005;  
PR 16-JUL-1999; 99US-0144085;  
PR 16-JUL-1999; 99US-0144086;  
PR 19-JUL-1999; 99US-0144325;  
PR 19-JUL-1999; 99US-0144331;  
PR 19-JUL-1999; 99US-0144332;  
PR 19-JUL-1999; 99US-0144333;  
PR 19-JUL-1999; 99US-0144334;  
PR 19-JUL-1999; 99US-0144335;  
PR 20-JUL-1999; 99US-0144352;  
PR 20-JUL-1999; 99US-0144632;  
PR 20-JUL-1999; 99US-0144884;  
PR 21-JUL-1999; 99US-0144814;  
PR 21-JUL-1999; 99US-0145086;  
PR 21-JUL-1999; 99US-0145088;  
PR 22-JUL-1999; 99US-0145085;  
PR 22-JUL-1999; 99US-0145087;  
PR 22-JUL-1999; 99US-0145089;  
PR 22-JUL-1999; 99US-0145192;  
PR 23-JUL-1999; 99US-0145145;  
PR 23-JUL-1999; 99US-0145218;  
PR 23-JUL-1999; 99US-0145224;  
PR 26-JUL-1999; 99US-0145276;  
PR 27-JUL-1999; 99US-0145913;  
PR 27-JUL-1999; 99US-0145918;  
PR 27-JUL-1999; 99US-0145919;  
PR 28-JUL-1999; 99US-0145921;  
PR 02-AUG-1999; 99US-0146386;  
PR 02-AUG-1999; 99US-0146388;  
PR 03-AUG-1999; 99US-0146389;  
PR 03-AUG-1999; 99US-0147038;  
PR 04-AUG-1999; 99US-0147284;  
PR 04-AUG-1999; 99US-0147302;  
PR 05-AUG-1999; 99US-0147192;  
PR 05-AUG-1999; 99US-0147260;  
PR 06-AUG-1999; 99US-0147303;  
PR 06-AUG-1999; 99US-0147416;  
PR 09-AUG-1999; 99US-0147493;  
PR 09-AUG-1999; 99US-0147935;  
PR 10-AUG-1999; 99US-0148171;  
PR 11-AUG-1999; 99US-0148319;  
PR 12-AUG-1999; 99US-0148341;  
PR 13-AUG-1999; 99US-0148565;  
PR 13-AUG-1999; 99US-0148684;  
PR 16-AUG-1999; 99US-0149368;  
PR 17-AUG-1999; 99US-0149175;  
PR 18-AUG-1999; 99US-0149426;  
PR 20-AUG-1999; 99US-0149722;  
PR 20-AUG-1999; 99US-0149723;  
PR 20-AUG-1999; 99US-0149929;  
PR 23-AUG-1999; 99US-0149902;  
PR 23-AUG-1999; 99US-0149930;  
PR 25-AUG-1999; 99US-0150566;  
PR 26-AUG-1999; 99US-0150894;  
PR 27-AUG-1999; 99US-0151065;  
PR 27-AUG-1999; 99US-0151066;  
PR 27-AUG-1999; 99US-0151080;  
PR 27-AUG-1999;

PR 30-AUG-1999; 99US-0151303;  
PR 31-AUG-1999; 99US-0151438;  
PR 01-SEP-1999; 99US-0151930;  
PR 07-SEP-1999; 99US-0152363;  
PR 10-SEP-1999; 99US-0153070;  
PR 13-SEP-1999; 99US-0153758;  
PR 15-SEP-1999; 99US-0154018;  
PR 16-SEP-1999; 99US-0154039;  
PR 20-SEP-1999; 99US-0154779;  
PR 22-SEP-1999; 99US-0155139;  
PR 23-SEP-1999; 99US-0155486;  
PR 24-SEP-1999; 99US-0155659;  
PR 28-SEP-1999; 99US-0156458;  
PR 29-SEP-1999; 99US-0156596;  
PR 04-OCT-1999; 99US-0157117;  
PR 05-OCT-1999; 99US-0157753;  
PR 06-OCT-1999; 99US-0157865;  
PR 07-OCT-1999; 99US-0158029;  
PR 08-OCT-1999; 99US-0158232;  
PR 12-OCT-1999; 99US-0158369;  
PR 13-OCT-1999; 99US-0159293;  
PR 13-OCT-1999; 99US-0159294;  
PR 14-OCT-1999; 99US-0159295;  
PR 14-OCT-1999; 99US-0159329;  
PR 14-OCT-1999; 99US-0159330;  
PR 14-OCT-1999; 99US-0159331;  
PR 14-OCT-1999; 99US-0159637;  
PR 14-OCT-1999; 99US-0159638;  
PR 18-OCT-1999; 99US-0159584;  
PR 21-OCT-1999; 99US-0160741;  
PR 21-OCT-1999; 99US-0160767;  
PR 21-OCT-1999; 99US-0160768;  
PR 21-OCT-1999; 99US-0160770;  
PR 21-OCT-1999; 99US-0160814;  
PR 21-OCT-1999; 99US-0160815;  
PR 22-OCT-1999; 99US-0160980;  
PR 22-OCT-1999; 99US-0160981;  
PR 22-OCT-1999; 99US-0160989;  
PR 25-OCT-1999; 99US-0161404;  
PR 25-OCT-1999; 99US-0161405;  
PR 25-OCT-1999; 99US-0161406;  
PR 26-OCT-1999; 99US-0161359;  
PR 26-OCT-1999; 99US-0161360;  
PR 26-OCT-1999; 99US-0161361;  
PR 28-OCT-1999; 99US-0161920;  
PR 28-OCT-1999; 99US-0161992;  
PR 28-OCT-1999; 99US-0161993;  
PR 29-OCT-1999; 99US-0162142;  
PR 29-OCT-1999;

Query Match 4.1%; Score 15; DB 21; Length 335;  
Best Local Similarity 100.0%; Pred. No. 4.1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 NNPERAVEVLYSGIP 188  
Db 134 NNPERAVEVLYSGIP 148  
|||||

## RESULT 13

AAG45205  
ID AAG45205 standard; Protein; 335 AA.

XX AAG45205;

AC AAG45205;

XX 18-OCT-2000 (first entry)

DT Arabidopsis thaliana protein fragment SEQ ID NO: 56724.

DE Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
KW Arabidopsis thaliana.  
XX OS

XX  
PN EF1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130049.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134368.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.

```
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159299.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.1%; Score 15; DB 21; Length 335;
Best Local Similarity 100.0%; Pred. No. 4.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 174 NNPERAVEYLYSGIP 188
Db 134 NNPERAVEYLYSGIP 148

RESULT 14
AAG44342
ID AAG44342 standard; Protein; 345 AA.
AC AAG44342;
XX
XX
XX
XX
XX 18-OCT-2000 (first entry)
DE
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55532.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
```

```
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
```

PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 4.1%; Score 15; DB 21; Length 345;  
Best Local Similarity 100.0%; Pred. No. 4.2e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 CDRNEELAANYLLEH 363  
DB 325 CDRNEELAANYLLEH 339  
|||||

RESULT 15  
AAG44341

ID AAG44341 standard; Protein; 368 AA.

XX AAG44341;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55531.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137503.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0155659.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.



```
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161520.
PR 28-OCT-1999; 99US-0161592.
PR 28-OCT-1999; 99US-0161593.
PR 29-OCT-1999; 99US-0162142.

Query Match      4.1%; Score 15; DB 21; Length 368;
Best Local Similarity 100.0%; Pred. No. 4.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 CDNEELAAVYLLEH 363
Db 348 CDNEELAAVYLLEH 362
|||||

RESULT 16
AAG19974
ID AAG19974 standard; Protein; 378 AA.
XX AC AAG19974;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21981.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 23-MAR-1999; 99US-0126264.
XX PR 23-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
```

PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155133.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158233.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 4.1%; Score 15; DB 21; Length 378;

Best Local Similarity 100.0%; Pred. No. 4.6e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 174 NNPERAVEYLYSGIP 188  
|||||  
Db 177 NNPERAVEYLYSGIP 191

#### RESULT 17

AAG45204

ID AAG45204 standard; Protein; 378 AA.

AC AAG45204;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 56723.

XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158212.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 22-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 4.1%; Score 15; DB 21; Length 378;

Best Local Similarity 100.0%; Pred. No. 4.6e-06; Mismatches 0; Indels 0; Gaps 0;

Qy 174 NNPERAVEYLYSGIP 188  
 |||||  
 Db 177 NNPERAVEYLYSGIP 191

## RESULT 18

AAV71458  
 ID AAY71458 standard; Protein; 405 AA.

AC AAV71458;  
 XX  
 DT 04-OCT-2000 (first entry)

DE Maize Rad23 protein #1.

XX Rad23; maize; ATCC No: PTA-530; recombinant expression cassette; wheat;  
 KW transgenic plant; soybean; sunflower, sorghum; canola; modulator.  
 XX

OS Zea mays.

PN WO200031268-A1.

XX 02-JUN-2000.

PF 12-OCT-1999; 99WO-US94129.

XX 23-NOV-1998; 98US-0109728.

XX (PION-) PIONEER HI-BRED INT INC.

PA Mahajan PB, Tagliani L;

PI WPI; 2000-400078/34.

DR N-PSDB; RAD01230.

XX Isolated nucleic acid encoding maize RAD23 protein is used to modulate  
 PT the levels of polypeptides in plant or in assays for identifying  
 PT compounds that bind to and/or increase/decrease enzymatic activity of  
 PT catalytically active polypeptides -

PS Claim 11b; Page 75-76; 82pp; English.

XX The present sequence is the maize Rad23 protein #1. It is isolated from  
 CC V5 root tissue of a Zea mays cell line B73, infested with corn root  
 CC worm. The cDNA sequence is deposited under the ATCC No: PTA-530. Maize  
 CC Rad23 DNA sequence operably linked to a promoter can be used to construct  
 CC a recombinant expression cassette. This expression cassette can be used  
 CC to generate a dicot or monocot transgenic plant e.g., maize, soybean,  
 CC sunflower, sorghum, canola, wheat, etc.. It can also be used to modulate  
 CC the levels of Rad23 polypeptide expression in a plant or in assays to  
 CC identify compounds, that bind to and/or modulate the enzymatic activity  
 CC of catalytically active polypeptides.

XX Sequence 405 AA;

Query Match 3.5%; Score 13; DB 21; Length 405;  
 Best Local Similarity 100.0%; Pred. No. 0.00048;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 167 RALRAAYNNPERA 179  
 |||||  
 Db 182 RALRAAYNNPERA 194

## RESULT 19

AAG33942  
 ID AAG33942 standard; Protein; 136 AA.

XX AAG33942;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 41222.

XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 XX termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.  
 PN  
 XX  
 PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 21-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134321.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 22-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142877.  
PR 13-JUL-1999; 99US-0143342.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 08-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3.0%; Score 11; DB 21; Length 136;  
Best Local Similarity 100.0%; Pred. No. 0.017;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 VKTLKGTHFEI 15  
|||||  
DB 5 VKTLKGTHFEI 15



XX WO200142279-A2.  
XX 14-JUN-2001.  
XX 13-DEC-2000; 2000WO-GB04781.  
XX 13-DEC-1999; 99GB-0029469.  
XX (PROT-) PROTEOM LTD.  
XX Roberts GW, Heal JR;  
XX WPI; 2001-381629/40.  
XX A set of peptide ligands for agricultural research and development of  
XX therapeutic agents comprise specific complementary peptides to proteins  
XX encoded by genes of plant genomes -  
XX Example 4; Page 31; 201pp; English.  
XX The present invention relates to a set of peptide ligands consisting of  
XX specific complementary peptides to proteins encoded by genes of plant  
XX genomes. The present sequence is one such peptide from Arabidopsis  
XX thaliana. The peptides of the present invention are useful in an assay to  
XX identify a peptide, especially a peptide pesticide or herbicide. The  
XX peptides are also useful for tools for agricultural research and  
XX development.  
XX Sequence 10 AA;  
XX Query Match 2.2%; Score 8; DB 22; Length 10;  
XX Best Local Similarity 100.0%; Pred. No. 1.3;  
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX A set of peptide ligands for agricultural research and development of  
XX therapeutic agents comprise specific complementary peptides to proteins  
XX encoded by genes of plant genomes -  
XX Example 4; Page 31; 201pp; English.  
XX The present invention relates to a set of peptide ligands consisting of  
XX specific complementary peptides to proteins encoded by genes of plant  
XX genomes. The present sequence is one such peptide from Arabidopsis  
XX thaliana. The peptides of the present invention are useful in an assay to  
XX identify a peptide, especially a peptide pesticide or herbicide. The  
XX peptides are also useful for tools for agricultural research and  
XX development.  
XX Sequence 10 AA;  
XX Query Match 2.2%; Score 8; DB 22; Length 10;  
XX Best Local Similarity 100.0%; Pred. No. 1.3;  
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX A set of peptide ligands for agricultural research and development of  
XX therapeutic agents comprise specific complementary peptides to proteins  
XX encoded by genes of plant genomes -  
XX Example 4; Page 32; 201pp; English.

XX The present invention relates to a set of peptide ligands consisting of  
XX specific complementary peptides to proteins encoded by genes of plant  
XX genomes. The present sequence is one such peptide from Arabidopsis  
XX thaliana. The peptides of the present invention are useful in an assay to  
XX identify a peptide, especially a peptide pesticide or herbicide. The  
XX peptides are also useful for tools for agricultural research and  
XX development.  
XX Sequence 10 AA;  
XX Query Match 2.2%; Score 8; DB 22; Length 10;  
XX Best Local Similarity 100.0%; Pred. No. 1.3;  
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 234 AGGGAGGG 241  
XX |||||  
XX 1 AGGGAGGG 8  
XX  
XX RESULT 24  
XX ABP34606  
XX ID ABP34606 standard; Protein; 60 AA.  
XX AC ABP34606;  
XX XX 08-JUL-2002 (first entry)  
XX DT 08-JUL-2002 (first entry)  
XX DE Human ORF3579 protein, SEQ ID NO:7158.  
XX KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
XX disease monitoring; cytokine; cell proliferation; cell differentiation;  
XX immune modulation; haematopoiesis regulation; tissue growth;  
XX angiogenesis; activin; inhibitor; chemotactic; chemokinetic; haemostatic;  
XX thrombolytic; tumour inhibition; bodily characteristic; fertility;  
XX behaviour; cancer; proliferative disorder; neurological disorder;  
XX cardiovascular disease; immune system disorder; organ transplantation;  
XX tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
XX hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
XX vasotropic; antipeptidic; antidiabetic; cytosolic; neurotropic;  
XX neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
XX cardiatic; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
XX dermatological; analgesic; virucide; antibacterial; fungicide.  
XX Homo sapiens.  
XX OS WO200190366-A2.  
XX PN 29-NOV-2001.  
XX PD 24-MAY-2001; 2001WO-US17076.  
XX PF 24-MAY-2000; 2000US-206690P.  
XX PR (CURA-) CURAGEN CORP.  
XX PA Leach WD, Shinkets RA;  
XX PI WPI; 2002-106200/14.  
XX DR N-PSDB; ABN78632.  
XX XX Novel human polypeptides and polynucleotides useful for diagnosing,  
XX preventing and treating cardiovascular disease, neurodegenerative,  
XX hyperproliferative disorders and disorders related to organ  
XX transplantation -  
XX Claim 10; Page 2036; 2508pp; English.  
XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
XX designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
XX ABN79587 represent cDNAs encoding them. The invention also encompasses  
XX polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
XX referred to as ORFX) proteins, polynucleotides at least 85% identical to

CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antinflammatory activity, tumour inhibition activity,  
 CC and anti-infective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.

XX Sequence 60 AA;

Query Match 2.2%; Score 8; DB 23; Length 60;

Best Local Similarity 100.0%; Pred. No. 7.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 TPATRQAP 100

|||||||

DB 21 TPATRQAP 28

RESULT 25

AAG39747

ID AAG39747 standard; Protein; 174 AA.

XX AC AAG39747;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 49227.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 17-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139452.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.



PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0147260.  
PR 02-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 15-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.2%; Score 8; DB 21; Length 174;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||  
DB 85 AGGAGGG 92

## RESULT 26

AAW79128  
ID AAW79128 standard; Protein; 265 AA.

XX AAW79128;

XX 18-NOV-1998 (first entry)

DT FLGA insert stabilising polypeptide.

DE Fusion protein; stabilising polypeptide; proteolytic degradation;  
resistance; half-life; autoimmune disease; inflammation; nitro drug;  
IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;  
nitroreductase protein; enzyme therapy; prodrug therapy; protease;  
cancer; pathological condition.

XX Epstein-barr virus.

OS WO9822577-A1.

PN 28-MAY-1998.

PD 17-NOV-1997; 97WO-IB01508.

PF 25-JUN-1997; 97US-0048945.

XX 15-NOV-1996; 96US-0030986.

XX (MASU/) MASUCCI M G.

XX Masucci MG;

XX WPI; 1998-312463/27.

DR N-PSDB; AAW55830.

XX New fusion proteins resistant to proteolytic degradation -  
comprising a core protein with a stabilising polypeptide comprising  
a peptide sequence containing glycine repeats  
PT  
XX  
PS Disclosure; Fig 3; 120pp; English.

XX This represents a stabilising polypeptide and is the FLGA insert of the  
 CC invention. The invention provides a method for increasing the resistance  
 CC of a core protein to proteolytic degradation that comprises linking or  
 CC inserting onto or into the core protein a stabilising polypeptide of  
 CC formula I (Glya)X(Glyb)Y(Glyc)Z<sub>n</sub> where Glya, Glyb, Glyc are 1-6  
 CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,  
 CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not  
 CC be identical from n repeat to n repeat. Alternatively a nucleic acid  
 CC encoding the stabilising polypeptide can be linked onto or inserted into  
 CC a nucleic acid encoding a core protein. The fusion proteins of the  
 CC invention are more resistant to degradation by proteases and, thus, have  
 CC a longer half-life than the unfused core protein. The products can be  
 CC used for treating autoimmune diseases, cancer and inflammation. In  
 CC particular, the core protein may be an IkappaB regulator protein for the  
 CC treatment of inflammatory bowel disease, or a nitroreductase protein  
 CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer  
 CC or other pathological conditions. The fusion proteins can also be used in  
 CC diagnostic methods such as in vivo imaging.

SQ Sequence 265 AA;

Query Match 2.2%; Score 8; DB 19; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
 |||||  
 Db 78 AGGGAGGG 85

RESULT 27

AAAP93704  
 ID AAP93704 standard; Protein; 325 AA.

XX AC AAP93704;

XX DT 25-MAR-2003 (updated)

XX DT 15-MAR-1992 (first entry)

XX DE Sequence of the 65kd surface antigen recognised by monoclonal  
 XX DE antibodies 7D1, 7D4 and 20C6.

XX KW Vaccine; coccidiosis; poultry; bivalent vaccine.

XX OS Eimeria.

XX PN EP344808-A.

XX PD 06-DEC-1989.

XX PF 02-JUN-1980; 80EP-0110056.

XX PR 03-JUN-1988; 88US-0202721.

XX PA (HOFF ) HOFFMANN-LA ROCHE AG.

XX PI Altenburger W, Binger MH, Chizzonite RA, Kramer RA;

XX PI Lomedico PT, McAndrew SJ;

XX DR WPI; 1989-358220/49.

XX DR N-PSDB; AAN92577.

XX PT DNA sequences encoding Eimeria surface antigens - used in  
 XX PT pox:virus vectors as a vaccine to protect chicks against  
 XX PT coccidiosis

XX PS Claim 3; Fig 17; 78pp; English.

XX CC The inventors claim a new protein which comprises one or more  
 CC immunoreactive and/or antigenic determinants of an Eimeria surface  
 CC antigen of mol. wt. 28,37,120 or more than 200 kD which specifically  
 CC binds to one or more monoclonal antibody (MaB) from ATCC HB 9707-9712

CC (see AAP93703-6). Also new are DNA encoding the protein (see  
 CC AAN92576-9), and a vaccine comprising one or more proteins. Vaccine  
 CC utility can be enhanced by inserting additional genes into the carrier  
 CC virus (see AAP91652).  
 CC (Updated on 25-MAR-2003 to correct PF field.)

SQ Sequence 325 AA;

Query Match 2.2%; Score 8; DB 10; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 NAGGGAGG 240

|||||  
 Db 225 NAGGGAGG 232

RESULT 28

AAAW31583  
 ID AAW31583 standard; Protein; 325 AA.

XX AC AAW31583;

XX DT 25-MAR-2003 (updated)

XX DT 21-MAY-1998 (first entry)

XX DE Eimeria tenella sporozoite, schizont, merozoite antigen.

XX KW Coccidiosis; vaccine; poultry; protozoan; parasite; antigen;

XX KW sporozoite; schizont; merozoite.

XX OS Eimeria tenella.

XX PH Key Location/Qualifiers

XX FT Peptide 1..20

XX FT /note= "the first approx. 20 amino acids of the  
 XX FT sequence may result from an artifact during  
 XX FT cDNA synthesis which caused a looping back  
 XX FT of a more internal sequence"

XX PN US5661015-A.

XX PD 26-AUG-1997.

XX PF 20-DEC-1991; 91US-0812349.

XX PR 20-DEC-1991; 91US-0812349.

XX PR 03-JUN-1988; 88US-0202721.

XX PA (HOFF ) HOFFMANN LA ROCHE INC.

XX PI Altenburger W, Binger M, Chizzonite RA, Kramer RA;

XX PI Lomedico PT, McAndrew SJ;

XX DR WPI; 1997-434379/40.

XX DR N-PSDB; AAT93594.

XX PT New DNA from Eimeria tenella and related immunogenic polypeptides -  
 XX PT useful in vaccines to protect poultry against coccidiosis

XX PS Example 6.5; Fig 17; 72pp; English.

XX CC This 65 kDa protein is recognised by monoclonal antibodies 7D1,  
 XX CC 20C6 and 7D4 (ATCC HB 9707). These antibodies also specifically  
 XX CC react with an Eimeria tenella 120 kDa surface antigen that is  
 XX CC present in the sporozoite, schizont and merozoite developmental  
 XX CC stages. The amino acid sequence was deduced from a cDNA clone  
 XX CC (see AAT93594) obtained from a cDNA library by immunological  
 XX CC screening with monoclonal antibodies raised against Eimeria  
 XX CC antigens. The invention provides DNA sequences (see AAT93593-98)  
 XX CC coding for Eimeria surface antigens (see AAW31582-84 and AAW33621-26),  
 XX CC recombinant vectors containing such DNA sequences, transformed  
 XX CC microorganisms containing such vectors, and methods for producing

CC the antigens using the transformed microorganisms. Methods are  
 CC also provided for protecting poultry against coccidiosis using the  
 CC Eimeria surface antigens. The surface antigens are administered  
 CC either as purified proteins or in the form of DNA encoding the  
 CC proteins in a viral vector such as a vaccinia virus. The vaccines  
 CC may produce antibodies that are cross-reactive with other Eimeria  
 CC species.  
 CC (Updated on 25-MAR-2003 to correct PF field.)  
 XX  
 XX  
 SQ Sequence 325 AA;  
 Query Match 2.2%; Score 8; DB 18; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 233 NAGGGAGG 240  
 Db 225 NAGGGAGG 232  
 |||||  
 RESULT 29  
 ABR01845  
 ID ABR01845 standard; Protein; 331 AA.  
 XX  
 AC ABR01845;  
 XX  
 DT 19-MAY-2003 (first entry)  
 XX  
 DE Human cancer-related protein, 161P2B7A v.3.  
 XX  
 KW Human; cytostatic; vaccine; cancer; immune response.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200283921-A2.  
 XX  
 PD 24-OCT-2002.  
 XX  
 PF 10-APR-2002; 2002WO-US11654.  
 XX  
 PR 10-APR-2001; 2001US-282739P.  
 PR 10-APR-2001; 2001US-283112P.  
 PR 25-APR-2001; 2001US-286630P.  
 XX  
 XX (AGEN-) AGENSYS INC.  
 XX  
 PI Jakobovits A, Challita-Rid PM, Paris M, Ge W, Hubert RS;  
 PI Morrison K, Morrison RK, Raitano AB;  
 XX  
 DR WPI; 2003-075555/07.  
 DR N-PSDB; ABZ78155.  
 XX  
 XX New composition comprising a substance that modulates the structure of  
 PT proteins and polynucleotides, useful for therapeutic, prognostic and  
 PT diagnostic reagents for eliciting cellular or humoral immune response  
 in cancer patients -  
 PT  
 XX  
 XX Example 5; Fig 12j; 1021pp; English.  
 PS  
 XX The present invention relates to novel human cancer-related genes and  
 CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and  
 CC proteins are useful for eliciting a humoral or cellular immune response.  
 CC The genes are useful as probes and primers for the amplification and/or  
 CC detection of genes, mRNAs or their fragments, as reagents for the  
 CC diagnosis and/or prognosis of cancer, as coding sequences capable of  
 CC directing the expression of the protein, as tools for modulating or  
 CC inhibiting the expression of genes and/or translation of transcripts, and  
 CC as therapeutic agents. The proteins and peptides are useful as  
 CC therapeutic, prognostic and diagnostic reagents for cancer.  
 XX  
 XX  
 SQ Sequence 331 AA;  
 Query Match 2.2%; Score 8; DB 24; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 234 AGGGAGGG 241  
 Db 83 AGGGAGGG 90  
 |||||  
 RESULT 30  
 AAB29250  
 ID AAB29250 standard; Protein; 332 AA.  
 XX  
 AC AAB29250;  
 XX  
 DT 07-FEB-2001 (first entry)  
 XX  
 DE Barley NADPH-thioredoxin reductase.  
 XX  
 KW Barley; NADPH-thioredoxin reductase; disulfide bridge reduction; NTR;  
 KW development; thioredoxin h; hair care product; venom neutralisation;  
 KW food technology; food allergy.  
 XX  
 OS Hordeum vulgare.  
 XX  
 PN WO200058352-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08566.  
 XX  
 PR 31-MAR-1999; 99US-0127198.  
 PR 06-DEC-1999; 99US-0169162.  
 PR 21-JAN-2000; 2000US-0177739.  
 PR 21-JAN-2000; 2000US-0177740.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Cho M, Del Val G, Caillau M, Lemaux PG, Buchanan BB;  
 XX  
 DR WPI; 2000-679291/66.  
 DR N-PSDB; AAC62460.  
 XX  
 PT Recombinant or isolated nucleic acid, useful for producing transgenic  
 PT plants with altered redox properties, encode thioredoxin h or  
 PT thioredoxin reductase -  
 XX  
 PS Claim 49; Fig 4; 125pp; English.  
 XX  
 CC The present invention relates to the isolation and use of the barley  
 CC thioredoxin h and NADPH-thioredoxin reductase coding and protein  
 CC sequences. Thioredoxin is thought to be involved in plant development via  
 CC its function in the reduction of disulfide bridges. Thioredoxin can be  
 CC used in hair care products and in the neutralisation of some venoms and  
 CC toxins, and is also useful in the reduction of some food, for example it  
 CC can be used to reduce the allergenicity of foods and the digestibility  
 CC of some proteins. It can also be used to enhance the baking qualities of  
 CC cereal flour.  
 XX  
 XX Sequence 332 AA;  
 SQ  
 Query Match 2.2%; Score 8; DB 21; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 236 GGAGGGGGL 243  
 Db 221 GGAGGGGGL 228  
 |||||  
 RESULT 31  
 AAR05217  
 ID AAR05217 standard; protein; 352 AA.  
 XX

AC AAR05217;  
 XX 25-MAR-2003 (updated)  
 DT 02-AUG-1990 (first entry)  
 XX  
 DE Antigen GX3271 encoded by cDNA of Eimeria tenella.  
 XX  
 KW Eimeria tenella; antigen GX3271; avian coccidiosis.  
 XX  
 OS Eimeria tenella.  
 XX  
 PN WO9000403-A.  
 XX  
 XX 25-JAN-1990.  
 PD  
 XX  
 PF 05-JUL-1989; 89WO-US02918.  
 XX  
 XX 05-JUL-1988; 88US-0215162.  
 PR  
 XX (GEMX ) GENEX CORP.  
 PA  
 XX Anderson DM, McCandless RJ, Strausberg SL, Strausberg RL;  
 PI  
 XX WPI; 1990-051586/07.  
 DR  
 XX N-PSDB; AAQ03319.  
 XX  
 XX Cloned gene or fragment encoding antigenic protein -  
 PT which binds with antibodies against avian coccidia, and  
 PT transformed cells used in vaccine  
 XX  
 PS Claim 5; Page 93; Fig 3; 134pp; English.  
 CC  
 CC An E. tenella cDNA library was screened with chicken immune serum. Two  
 CC phage plaques were identified that produce coccidial antigens reactive  
 CC with this chicken immune serum. The coccidial DNA from these plaques was  
 CC transferred to bacteriophage M13 for DNA sequence analysis. Their DNA  
 CC sequences do not show homology with each other, nor with the GX5401  
 CC (AA93324) coding sequence. Plasmid expression vectors were constructed  
 CC for the production of them in E. coli. They were designated PGX3271 and  
 CC PGX3273, and encode beta-galactosidase coccidial Ag. fusion proteins.  
 CC Also new are host cells transformed with the vectors. The transformed  
 CC cells are used in vaccine to immunise birds against avian coccidiosis.  
 CC By labelling the peptides, they can be used as a type-specific probe.  
 CC The AP may also be used in an assay to detect Ab against the coccidia.  
 CC The Abs are used to identify transformed cells contg. the DNA.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 352 AA;  
 Query Match 2.2%; Score 8; DB 11; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 233 NAGGGAGG 240  
 Db 252 NAGGGAGG 259  
 |||||  
 |||||  
 RESULT 32  
 ABB65875  
 ID ABB65875 standard; Protein; 566 AA.  
 XX  
 AC ABB65875;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 24417.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PI

PN WO200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR  
 XX 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 XX WPI; 2001-656860/75.  
 DR  
 XX N-PSDB; ABL09978.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 24417; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABBS7737-ABBS72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 566 AA;  
 Query Match 2.2%; Score 8; DB 22; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 234 AGGGAGGG 241  
 Db 95 AGGGAGGG 102  
 |||||  
 |||||  
 RESULT 33  
 AAY57087  
 ID AAY57087 standard; protein; 604 AA.  
 XX  
 AC AAY57087;  
 XX  
 DT 28-FEB-2000 (first entry)  
 XX  
 DE Mouse retinoid binding protein.  
 XX  
 KW Neglected target tissue antigen; NNTA; autoimmunity; autoimmune response;  
 KW immunotherapeutic agent; insulin dependent diabetes mellitus;  
 KW multiple sclerosis; autoimmune thyroiditis; rheumatoid arthritis;  
 KW uveoretinitis; inflammatory response.  
 XX  
 OS Mus sp.  
 XX  
 PN WO9956763-A1.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 07-MAY-1999; 99WO-US10250.  
 XX  
 XX 07-MAY-1998; 98US-0084636.  
 PR  
 XX (REGC ) UNIV CALIFORNIA.  
 PA  
 XX Kaufman DL, Tian J, Olcott A;  
 PI

XX WPI; 2000-052905/04.  
 XX Administration of neglected target tissue antigens to modulate immune  
 PT responses -  
 PT  
 XX Disclosure; Page 29; 79pp; English.  
 XX  
 CC Amino acid sequences AAY57063-Y57091 are examples of neglected target  
 CC tissue antigens NTAs. NTAs are antigens (whole antigens or fragments)  
 CC not involved in autoimmunity. These peptides and proteins are used in  
 CC the method of the invention which involves administering an NTA as an  
 CC antigen based immunotherapeutic agent, to a host afflicted with an  
 CC autoimmune response associated with an autoimmune disease. The  
 CC immunotherapeutic agent is used to treat autoimmune diseases such as  
 CC insulin dependent diabetes mellitus, multiple sclerosis, autoimmune  
 CC thyroiditis, uveoretinitis, rheumatoid arthritis or abnormal  
 CC inflammatory immune responses. The NTA induces regulatory tolerance by  
 CC elicitation of regulatory T cells among T cells recognizing the NTA but  
 CC not participating in the immune response. The NTA are capable of  
 CC recognition by substantial populations of uncommitted T cells which can  
 CC be primed, or biased, towards regulatory responses to provide effective  
 CC treatment. The NTA are effective in regulating undesirable immune  
 CC responses even when target determinants used as agents promoting  
 CC tolerance agents have failed to induce an effective regulatory T cell  
 CC response. NTAs as agents promoting tolerance are anticipated to be safer  
 CC than use of target determinants.  
 XX  
 SQ Sequence 604 AA;  
 Query Match 2.2%; Score 8; DB 21; Length 604;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 102 LEAPQAP 109  
 DB 90 LEAPQAP 97  
 RESULT 34  
 ABB69938  
 ID ABB69938 standard; Protein; 620 AA.  
 XX  
 AC ABB69938;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 36606.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li FWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABLJ4041.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

XX Disclosure; SEQ ID NO 36606; 21pp + Sequence Listing; English.  
 PS  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 620 AA;  
 Query Match 2.2%; Score 8; DB 22; Length 620;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 234 AGGGAGGG 241  
 DB 26 AGGGAGGG 33  
 RESULT 35  
 AAR07094  
 ID AAR07094 standard; protein; 712 AA.  
 XX  
 AC AAR07094;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 24-JAN-1991 (first entry)  
 XX  
 DE Protein having immunological properties of E.tenella.  
 XX  
 KW Coccidiosis; vaccine; monoclonal antibodies.  
 XX  
 OS Eimeria tenella.  
 XX  
 PN EP390267-A.  
 XX  
 PD 03-OCT-1990.  
 XX  
 PF 22-MAR-1990; 90EP-0200686.  
 XX  
 PR 22-MAR-1990; 90EP-0200686.  
 XX  
 PA (ALKU ) AKZO NV.  
 PA (ALKU ) AKZO NOBEL NV.  
 XX  
 PI Dijkema R, Vermeulen A, Clarke LE, Tomly FM;  
 XX  
 DR WPI; 1990-299216/40.  
 DR N-PSDB; AAQ06033.  
 XX  
 PT Coccidiosis vaccine protein - obtained by immune adsorption of  
 PT eimeria tenella extract to substrate containing monoclonal  
 PT antibody and releasing adsorption fraction  
 XX  
 PS Disclosure; fig 6; 28pp; English.  
 XX  
 CC This E.tenella protein can be used in bird vaccines for protection  
 CC against Coccidiosis. Constituent polypeptides can also be used,  
 CC or antibodies raised against them. See also AAR07092-93.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 712 AA;  
 Query Match 2.2%; Score 8; DB 11; Length 712;  
 Best Local Similarity 100.0%; Pred. No. 77;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NAGGAGG 240  
 |||||  
 Db 612 NAGGAGG 619

RESULT 36  
 AAW56163  
 ID AAW56163 standard; Protein; 738 AA.  
 XX  
 AC AAW56163;  
 XX  
 DT 28-JUL-1998 (first entry)  
 XX  
 DE New DNA sequence isolated from Pinctada fucata.  
 XX  
 KW Pinctada fucata; protein; cosmetic.  
 XX  
 OS Pinctada fucata.  
 XX  
 PN JP10080285-A.  
 XX  
 PD 31-MAR-1998.  
 XX  
 PF 28-MAY-1997; 97JP-0138461.  
 XX  
 PR 15-JUL-1996; 96JP-0184459.  
 XX  
 PA (MIKI-) MIKIMOTO SEIYAKU KK.  
 XX  
 XX WPI; 1998-254410/23.  
 DR N-PSDB; AAV22683.  
 XX  
 PT New cDNA and e.g. vector, host cell and polypeptide - used to  
 PT produce polypeptide in high yields, which is used in cosmetics  
 PS Claim 9; Pages 9-11; 15pp; Japanese.  
 XX  
 CC The present sequence represents protein encoded by a new DNA sequence  
 CC isolated from Pinctada fucata. The protein be used as an ingredient  
 CC in cosmetics.  
 XX  
 SQ Sequence 738 AA;

Query Match 2.2%; Score 8; DB 19; Length 738;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGG 241  
 |||||  
 Db 119 AGGGAGG 126

RESULT 37  
 ABP69286  
 ID ABP69286 standard; Protein; 828 AA.  
 XX  
 AC ABP69286;  
 XX  
 DT 20-JAN-2003 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 1333.  
 XX  
 KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 KW haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritic.

OS Homo sapiens.  
 XX  
 PN WO200270539-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 05-MAR-2002; 2002WO-US05095.  
 XX  
 PR 05-MAR-2001; 2001US-0799451.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX  
 DR WPI; 2002-759812/82.  
 DR N-PSDB; ABZ11503.  
 XX  
 PT New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
 PT platelet or coagulation disorders -  
 XX  
 PS Claim 9; SEQ ID NO 1333; 1012pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleotide sequence selected from any of 948 sequences  
 CC (ABZ11119-ABZ12066) or their mature protein coding portion, active domain  
 CC coding protein or complementary sequences. The polynucleotides are useful  
 CC for identifying expressed genes or for physical mapping of human genome.  
 CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
 CC weight markers, as a food supplement, for generating antibodies, in  
 CC medical imaging, screening and diagnostic assays and for treating  
 CC cell-proliferative disorders (cancer), neurodegenerative diseases  
 CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
 CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
 CC disorders, platelet or coagulation disorders, wound, burns, incision,  
 CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
 CC parasitic), arthritis, etc.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 828 AA;

Query Match 2.2%; Score 8; DB 23; Length 828;  
 Best Local Similarity 100.0%; Pred. No. 89;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGG 241  
 |||||  
 Db 706 AGGGAGG 713

RESULT 38  
 ABP58225  
 ID ABP58225 standard; Protein; 828 AA.  
 XX  
 AC ABP58225;  
 XX  
 DT 31-MAR-2003 (first entry)  
 XX  
 DE Human cell adhesion and extracellular matrix protein 2.  
 XX  
 KW Cell adhesion and extracellular matrix protein 2; CADECM-2; human;  
 KW anti-HIV; virucide; anti-allergic; anti-inflammatory; antianemic;  
 KW antiparkinsonian; nootropic; anticonvulsant; antiinfertility;  
 KW antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;  
 KW cyostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;  
 KW angiot; thymimetic; neuroprotective; osteopathic; antiarthritic;  
 KW antiparasitic; antihelminthic; antipsoriatic; uropathic;  
 KW ophthalmological; antirheumatic; haemostatic; antibacterial;

KW protozoacide; fungicide; gynaecological; cadherin; gene therapy.  
XX Homo sapiens.  
OS  
XX WO200288322-A2.  
PN  
XX  
XX  
XX 07-NOV-2002.  
PD  
XX  
XX 01-MAY-2002; 2002WO-US13874.  
PF  
XX  
XX 02-MAY-2001; 2001US-288290P.  
PR 21-MAY-2001; 2001US-292468P.  
PR 15-JUN-2001; 2001US-298616P.  
PR 28-JUN-2001; 2001US-301672P.  
PR 04-JAN-2002; 2002US-345008P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
PA  
XX Yue H, Lee EA, Duggan BM, Thangavelu K, Honchell CD, Ding L;  
PI Hillman JL, Baughn MR, Kallick DA, Lee S, Warren BA, Xu Y;  
PI Tran UK, Lal PG, Thornton M, Hafalia AJA, Yao MG, Nguyen DB;  
PI Gandhi AR, Khan FA, Walhia NK, Griffin JA, Chinn AM, Elliott VS;  
PI Ramkumar J, Arvizu CS, Forsythe IJ;  
XX WPI; 2003-167112/16.  
DR N-PSDB; AB224579.  
XX  
XX New human cell adhesion and extracellular matrix proteins, useful for  
PT diagnosing, treating or preventing autoimmune or inflammatory disorder  
PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
PT cancer or hepatitis -  
XX  
XX Claim 1; Page 140-141; 178pp; English.  
PS  
XX  
XX The present sequence is the protein sequence of Incyte polypeptide  
CC 141780CD1 denoted human cell adhesion and extracellular matrix  
CC protein 2 (CADECM-2). The protein is encoded by a clone isolated  
CC from a male cerebellum cDNA library. Homology searches indicate  
CC it to be a cadherin. The invention provides CADECM-1 to -11  
CC polypeptides (see ABP58224-34) and polynucleotides (see AB224578-88),  
CC expression vectors, host cells, antibodies, agonists and antagonists.  
CC These are useful for diagnosing, treating or preventing disorders  
CC associated with aberrant expression of CADECM, particularly cell  
CC proliferative disorders (e.g. arteriosclerosis, atherosclerosis,  
CC cirrhosis, hepatitis, paroxysmal nocturnal haemoglobinuria,  
CC polycythaemia vera, psoriasis, primary thrombocytopaenia or  
CC cancer), developmental disorders (e.g. renal tubular acidosis,  
CC anaemia or mental retardation), neurological disorders (e.g.  
CC Alzheimer's disease, Parkinson's disease or epilepsy), reproductive  
CC disorders (e.g. infertility or a disruption in the menstrual cycle),  
CC or autoimmune/inflammatory disorders (e.g. AIDS, allergy, asthma,  
CC autoimmune thyroiditis, contact dermatitis, Crohn's disease,  
CC diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,  
CC Graves' disease, Hashimoto's thyroiditis, irritable bowel syndrome,  
CC multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis,  
CC Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome,  
CC uveitis), or viral, bacterial, fungal, parasitic, protozoal or  
CC helminthic infections.  
XX  
XX Sequence 828 AA;  
SQ  
  
Query Match 2.2%; Score 8; DB 24; Length 828;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 234 AGGGAGGG 241  
| | | | |  
Db 706 AGGGAGGG 713  
  
RESULT 39  
AAG91922  
ID AAG91922 standard; Protein; 1004 AA.

XX AAG91922;  
AC  
XX 26-SEP-2001 (first entry)  
DT  
XX C glutamic protein fragment SEQ ID NO: 5676.  
DE  
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis.  
XX  
XX Corynebacterium glutamicum.  
OS  
XX EP1108790-A2.  
PN  
XX 20-JUN-2001.  
PD  
XX  
XX 18-DEC-2000; 2000EP-0127688.  
PF  
XX  
XX 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
PA  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
PI  
XX WPI; 2001-376931/40.  
DR N-PSDB; AAH67141.  
XX  
XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
XX Claim 17; SEQ ID NO: 5676; 246pp + Sequence Listing; English.  
PS  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 1004 AA;  
  
Query Match 2.2%; Score 8; DB 22; Length 1004;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 234 AGGGAGGG 241  
| | | | |  
Db 338 AGGGAGGG 345  
  
RESULT 40  
ABP71196  
ID ABP71196 standard; Protein; 1088 AA.  
XX  
XX AC ABP71196;  
XX  
XX 14-APR-2003 (first entry)  
DT  
XX S. cinamomeus cinnamycin cinM protein.  
DE  
XX Cinnamycin; bacterium; cinA; cinM; cinX; cinT; cinH; cinY; antibiotic;  
KW

KW food additive; antibacterial.  
 XX  
 OS Streptomyces cinamomeus.  
 XX  
 PN WO200288367-A1.  
 XX  
 XX WO200288367-A1.  
 PD  
 XX 07-NOV-2002.  
 XX  
 XX 29-APR-2002; 2002WO-GB01983.  
 PF  
 XX 27-APR-2001; 2001GB-0010432.  
 PR  
 XX (PLAN-) PLANT BIOSCIENCE LTD.  
 PA (WIDD/) WIDDICK D A.  
 XX  
 XX Bibb MJ;  
 PI  
 XX WPI: 2003-111893/10.  
 DR N-PSDB; ABZ58812.  
 DR  
 XX New expression cassettes or genes isolated from Streptomyces  
 PT cinamomeus, useful for producing a library of lantibiotic-producing  
 PT host cells or lantibiotics, which are useful as food additives and  
 PT antibacterial agents -  
 XX  
 PS Claim 13; Fig 12; 11pp; English.  
 XX  
 CC The invention relates to expression cassettes or sets of nucleic acids  
 CC comprising various open reading frames selected from (a) a cina open  
 CC reading frame (orf), a cinM orf, and optionally a cinX orf; or (b) a cina  
 CC orf, a cinM orf, a cinT orf, a cinH orf, a cinY orf, and optionally a  
 CC cinX orf. The expression cassettes, set of nucleic acids, (set of)  
 CC vectors, or methods are useful for producing a library of lantibiotic-  
 CC producing host cells or a library of lantibiotics. These are particularly  
 CC useful for producing lantibiotic cinamycin or its modified versions. The  
 CC lantibiotics are useful as antibiotics having efficacy and utility as  
 CC food additives and antibacterial agents against Propionibacterium acnes  
 CC and problematic pathogens, e.g. methicillin-resistant Staphylococcus  
 CC aureus (which has or is developing resistance to many commonly used  
 CC antibiotics), or Streptococcus pneumoniae. Sequences ABP71191-71211  
 CC represent the various cinorf proteins encoded by the cinamycin cluster  
 CC from S. cinamomeus 40005 as present on the plasmid pDMF79.  
 XX  
 SQ Sequence 1088 AA;  
 Query Match 2.2%; Score 8; DB 24; Length 1088;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 234 AGGGAGGG 241  
 Db 1039 AGGGAGGG 1046  
 |||||  
 |||||  
 RESULT 41  
 AAU91279  
 ID AAU91279 standard; Protein; 1329 AA.  
 XX  
 AC AAU91279;  
 XX  
 XX 18-JUN-2002 (first entry)  
 DT  
 XX Human NOV3a protein.  
 DE  
 XX Human; NOVX; gene therapy; cardiomyopathy; atherosclerosis;  
 KW diabetes; cell signal processing; metabolic pathway modulation;  
 KW inflammation; autoimmune disorder; scleroderma; transplantation;  
 KW allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;  
 KW graft versus host disease; Leach-Nyhan syndrome; periodontitis;  
 KW pancreatitis; musculoskeletal disorder; Parkinson's disease;  
 KW Huntington's disease; behavioural disorder; pain; obesity; wound healing;  
 KW neurodegenerative disorder; neuropsychiatric disorder; hypertension;  
 KW growth disorder; reproductive disorder; lung disease.

XX Homo sapiens.  
 OS  
 XX WO200216600-A2.  
 PN  
 XX 28-FEB-2002.  
 PD  
 XX 27-AUG-2001; 2001WO-US26518.  
 PF  
 XX 25-AUG-2000; 2000US-227800P.  
 PR 25-AUG-2000; 2000US-228205P.  
 PR 25-AUG-2000; 2000US-228324P.  
 PR 30-AUG-2000; 2000US-228997P.  
 PR 30-AUG-2000; 2000US-229185P.  
 PR 01-SEP-2000; 2000US-229780P.  
 PR 01-SEP-2000; 2000US-229848P.  
 PR 01-SEP-2000; 2000US-229850P.  
 PR 22-JAN-2001; 2001US-263337P.  
 PR 31-JAN-2001; 2001US-265518P.  
 PR 15-MAR-2001; 2001US-276451P.  
 PR 27-MAR-2001; 2001US-279196P.  
 PR 24-AUG-2001; 2001US-0393398.  
 XX (CURA-) CURAGEN CORP.

Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K;  
 Spytek KA, Zerhusen BD, Rastelli L, Verney CAM, Patturajan M;  
 Tchernev VT, Padigar M, Taupier RJ;  
 DR  
 XX WPI: 2002-292064/33.  
 DR N-PSDB; ABK5563.

New isolated cytoplasmic, nuclear, membrane bound and secreted  
 polypeptides, termed NOVX, useful for treating inflammation, autoimmune  
 disorders, haemophilia, Leach-Nyhan syndrome, pancreatitis,  
 PT musculoskeletal disorders -

Claim 1; Page 32; 245pp; English.

The invention relates to an isolated cytoplasmic, nuclear, membrane bound  
 or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,  
 4, 5a, 5b, 5c, 5d, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a  
 mature form, or a variant of the mature form of NOVX. Also included  
 are a polynucleotide encoding NOVX (or its complement), a vector  
 comprising the polynucleotide, a cell comprising the vector, an  
 anti-NOVX antibody, determining the presence of NOVX in a sample  
 using the antibody, determining the presence of NOVX polynucleotide in  
 a sample using a probe which binds to NOVX polynucleotide, identifying a  
 an agent which binds to NOVX (including modulators of NOVX). NOVX, the  
 polynucleotide and the antibody are useful for diagnosing, treating  
 or preventing a NOVX-associated disorder selected from cardiomyopathy,  
 atherosclerosis, diabetes, a disorder related to cell signal processing,  
 and metabolic pathway modulation, inflammation, autoimmune disorders,  
 scleroderma, transplantation, allergies, systemic lupus erythematosus,  
 haemophilia, graft versus host disease, Alzheimer's disease, stroke,  
 Leach-Nyhan syndrome, periodontitis, pancreatitis, musculoskeletal  
 disorders, Parkinson's disease, Huntington's disease, behavioural  
 disorders, pain, neurodegenerative and neuropsychiatric disorders,  
 hypertension, wound healing, obesity, growth and reproductive  
 disorders, lung diseases and many other diseases and disorders listed in  
 the specification. NOVX, the polynucleotide and the antibody are useful  
 in screening assays, detection assays (e.g., chromosomal mapping, tissue  
 typing, forensic biology), predictive medicine (e.g., diagnostic assays,  
 prognostic assays, monitoring clinical trials and pharmacogenomic), and  
 in methods of treatment (e.g., therapeutic and prophylactic). NOVX is  
 useful as immunogen to produce antibodies immunospecific for NOVX, as  
 vaccines to screen for potential agonist and antagonist compounds; and  
 as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide  
 is useful in gene therapy, to express NOVX, to detect NOVX mRNA  
 or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The  
 vector is useful for producing non-human transgenic animals. The antibody  
 is useful for isolating, and purifying NOVX and to monitor protein levels  
 in tissue as part of a clinical testing procedure. The present sequence



CC represents a NOVX protein.

XX Sequence 1329 AA;

Query Match 2.2%; Score 8; DB 23; Length 1329;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241

|||||||

Db 745 AGGGAGGG 752

RESULT 42

ABB60138

ID ABB60138 standard; Protein; 1385 AA.

XX

AC ABB60138;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 7206.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

XX

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

XX

DR N-PSDB; ABL04241.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Disclosure; SEQ ID NO 7206; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 1385 AA;

Query Match

Best Local Similarity 2.2%; Score 8; DB 22; Length 1385;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241

|||||||

Db 344 AGGGAGGG 351

RESULT 43

ABB65356

ID ABB65356 standard; Protein; 1417 AA.

XX

AC ABB65356;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 22860.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

XX

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

XX

DR N-PSDB; ABL09459.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

PS Disclosure; SEQ ID NO 22860; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 1417 AA;

Query Match

Best Local Similarity 2.2%; Score 8; DB 22; Length 1417;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241

|||||||

Db 719 AGGGAGGG 726

RESULT 44

ABB61677

ID ABB61677 standard; Protein; 2075 AA.

XX

AC ABB61677;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 11823.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR N-PSDB; ABL05780.  
XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL05780.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 11823; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2075 AA;  
  
Query Match 2.2%; Score 8; DB 22; Length 2075;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 234 AGGGAGGG 241  
Db 143 AGGGAGGG 150  
|||||  
  
RESULT 45  
AAM00523  
ID AAM00523 standard; Peptide; 9 AA.  
XX  
AC AAM00523;  
XX  
DT 01-OCT-2001 (first entry)  
XX  
DE Human kinase fragment SEQ ID NO: 1071.  
XX  
XX Human; single nucleotide polymorphism; SNP; paternity test;  
KW forensic test; aberrant protein expression.  
XX  
OS Homo sapiens.  
XX  
PN WO200151670-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 05-JAN-2001; 2001WO-US00322.  
XX  
PR 07-JAN-2000; 2000US-0174962.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Leach MD;  
  
Query Match 1.9%; Score 7; DB 22; Length 9;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 236 GGAGGGP 242  
Db 1 GGAGGGP 7  
|||||  
  
RESULT 46  
AAM00524  
ID AAM00524 standard; Peptide; 9 AA.  
XX  
AC AAM00524;  
XX  
DT 01-OCT-2001 (first entry)  
XX  
DE Human kinase fragment SEQ ID NO: 1072.  
XX  
XX Human; single nucleotide polymorphism; SNP; paternity test;  
KW forensic test; aberrant protein expression.  
XX  
OS Homo sapiens.  
XX  
PN WO200151670-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 05-JAN-2001; 2001WO-US00322.  
XX  
PR 07-JAN-2000; 2000US-0174962.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Leach MD;  
XX  
DR WPI; 2001-451871/48.  
DR N-PSDB; AAM89641.  
XX  
XX Isolated human polynucleotides containing single nucleotide  
PT polymorphisms, useful for the treatment and diagnosis of e.g. cancer,  
PT infection and diabetes -  
XX  
XX Disclosure; Page 411; 475pp; English.  
XX  
XX The present invention relates to human nucleic acids containing single  
CC nucleotide polymorphisms (SNPs). These can be used in forensic and  
CC paternity tests, and to aid in the treatment of diseases associated with  
CC nucleotide polymorphisms (SNPs). These can be used in forensic and  
CC paternity tests, and to aid in the treatment of diseases associated with  
CC aberrant protein expression, including cancer, amyloidosis, diabetes,  
XX Shimkets RA, Leach MD;

XX WPI; 2001-451871/48.  
DR N-PSDB; AAM89640.  
XX  
PT Isolated human polynucleotides containing single nucleotide  
PT polymorphisms, useful for the treatment and diagnosis of e.g. cancer,  
PT infection and diabetes -  
XX  
XX Disclosure; Page 410; 475pp; English.  
XX  
XX The present invention relates to human nucleic acids containing single  
CC nucleotide polymorphisms (SNPs). These can be used in forensic and  
CC paternity tests, and to aid in the treatment of diseases associated with  
CC aberrant protein expression, including cancer, amyloidosis, diabetes,  
CC Alzheimer's disease, Down's syndrome, oedema, lupus (SLE), vasculitis,  
CC glomerulonephritis, haemolytic anaemia, thrombocytopenia, arthritis,  
CC meningitis, muscular disorders, dementia, neurological diseases, tuberculous  
CC sclerosis, male infertility, hypercalcaemia, blood pressure disorders,  
CC osteoporosis, pathogenic infections, hypercholesterolaemia, obesity or  
CC autoimmunity. The present sequence is a peptide encoded by a  
CC polymorphism-containing oligonucleotide fragment of the invention.  
XX  
XX Sequence 9 AA;  
  
Query Match 1.9%; Score 7; DB 22; Length 9;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 236 GGAGGGP 242  
Db 1 GGAGGGP 7  
|||||  
  
RESULT 46  
AAM00524  
ID AAM00524 standard; Peptide; 9 AA.  
XX  
AC AAM00524;  
XX  
DT 01-OCT-2001 (first entry)  
XX  
DE Human kinase fragment SEQ ID NO: 1072.  
XX  
XX Human; single nucleotide polymorphism; SNP; paternity test;  
KW forensic test; aberrant protein expression.  
XX  
OS Homo sapiens.  
XX  
PN WO200151670-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 05-JAN-2001; 2001WO-US00322.  
XX  
PR 07-JAN-2000; 2000US-0174962.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Leach MD;  
XX  
DR WPI; 2001-451871/48.  
DR N-PSDB; AAM89641.  
XX  
XX Isolated human polynucleotides containing single nucleotide  
PT polymorphisms, useful for the treatment and diagnosis of e.g. cancer,  
PT infection and diabetes -  
XX  
XX Disclosure; Page 411; 475pp; English.  
XX  
XX The present invention relates to human nucleic acids containing single  
CC nucleotide polymorphisms (SNPs). These can be used in forensic and  
CC paternity tests, and to aid in the treatment of diseases associated with  
CC aberrant protein expression, including cancer, amyloidosis, diabetes,

CC Alzheimer's disease, Down's syndrome, oedema, lupus (SLE), vasculitis,  
 CC glomerulonephritis, haemolytic anaemia, thrombocytopaenia, arthritis,  
 CC meningitis, muscular disorders, dementia, neurological diseases, tubercu-  
 CC sclerosis, male infertility, hypercalcaemia, blood pressure disorders,  
 CC osteoporosis, pathogenic infections, hypercholesterolaemia, obesity or  
 CC autoimmunity. The present sequence is a peptide encoded by a  
 CC polymorphism-containing oligonucleotide fragment of the invention.  
 XX  
 SQ Sequence 9 AA;

Query Match 1.9%; Score 7; DB 22; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0;

QY 236 GGAGGGP 242  
 DB 1 GGAGGGP 7  
 |||||

RESULT 47  
 AAM00521  
 ID AAM00521 standard; Peptide; 10 AA.  
 XX  
 AC AAM00521;  
 XX  
 DT 01-OCT-2001 (first entry)  
 XX  
 DE Human kinase fragment SEQ ID NO: 1069.  
 XX  
 KW Human; single nucleotide polymorphism; SNP; paternity test;  
 KW forensic test; aberrant protein expression.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200151670-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PF 05-JAN-2001; 2001WO-US00322.  
 PR 07-JAN-2000; 2000US-0174962.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 PI Shimkets RA, Leach MD;  
 DR WPI; 2001-451871/48.  
 DR N-PSDB; AAH89638.

XX Isolated human polynucleotides containing single nucleotide  
 PT polymorphisms, useful for the treatment and diagnosis of e.g. cancer,  
 PT infection and diabetes -  
 XX  
 PS Disclosure; Page 410; 475pp; English.  
 XX  
 CC The present invention relates to human nucleic acids containing single  
 CC nucleotide polymorphisms (SNPs). These can be used in forensic and  
 CC paternity tests, and to aid in the treatment of diseases associated with  
 CC aberrant protein expression, including cancer, amyloidosis, diabetes,  
 CC Alzheimer's disease, Down's syndrome, oedema, lupus (SLE), vasculitis,  
 CC glomerulonephritis, haemolytic anaemia, thrombocytopaenia, arthritis,  
 CC meningitis, muscular disorders, dementia, neurological diseases, tubercu-  
 CC sclerosis, male infertility, hypercalcaemia, blood pressure disorders,  
 CC osteoporosis, pathogenic infections, hypercholesterolaemia, obesity or  
 CC autoimmunity. The present sequence is a peptide encoded by a  
 CC polymorphism-containing oligonucleotide fragment of the invention.  
 XX  
 SQ Sequence 10 AA;

Query Match 1.9%; Score 7; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGP 242  
 DB 1 GGAGGGP 7  
 |||||

RESULT 48  
 AAM00522  
 ID AAM00522 standard; Peptide; 10 AA.  
 XX  
 AC AAM00522;  
 XX  
 DT 01-OCT-2001 (first entry)  
 XX  
 DE Human kinase fragment SEQ ID NO: 1070.  
 XX  
 KW Human; single nucleotide polymorphism; SNP; paternity test;  
 KW forensic test; aberrant protein expression.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200151670-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PF 05-JAN-2001; 2001WO-US00322.  
 PR 07-JAN-2000; 2000US-0174962.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 PI Shimkets RA, Leach MD;  
 DR WPI; 2001-451871/48.  
 DR N-PSDB; AAH89638.

QY 236 GGAGGGP 242  
 DB 2 GGAGGGP 8  
 |||||

RESULT 48  
 AAM00522  
 ID AAM00522 standard; Peptide; 10 AA.  
 XX  
 AC AAM00522;  
 XX  
 DT 01-OCT-2001 (first entry)  
 XX  
 DE Human kinase fragment SEQ ID NO: 1070.  
 XX  
 KW Human; single nucleotide polymorphism; SNP; paternity test;  
 KW forensic test; aberrant protein expression.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200151670-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PF 05-JAN-2001; 2001WO-US00322.  
 PR 07-JAN-2000; 2000US-0174962.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 PI Shimkets RA, Leach MD;  
 DR WPI; 2001-451871/48.  
 DR N-PSDB; AAH89638.

XX Isolated human polynucleotides containing single nucleotide  
 PT polymorphisms, useful for the treatment and diagnosis of e.g. cancer,  
 PT infection and diabetes -  
 XX  
 PS Disclosure; Page 410; 475pp; English.  
 XX  
 CC The present invention relates to human nucleic acids containing single  
 CC nucleotide polymorphisms (SNPs). These can be used in forensic and  
 CC paternity tests, and to aid in the treatment of diseases associated with  
 CC aberrant protein expression, including cancer, amyloidosis, diabetes,  
 CC Alzheimer's disease, Down's syndrome, oedema, lupus (SLE), vasculitis,  
 CC glomerulonephritis, haemolytic anaemia, thrombocytopaenia, arthritis,  
 CC meningitis, muscular disorders, dementia, neurological diseases, tubercu-  
 CC sclerosis, male infertility, hypercalcaemia, blood pressure disorders,  
 CC osteoporosis, pathogenic infections, hypercholesterolaemia, obesity or  
 CC autoimmunity. The present sequence is a peptide encoded by a  
 CC polymorphism-containing oligonucleotide fragment of the invention.  
 XX  
 SQ Sequence 10 AA;

Query Match 1.9%; Score 7; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGP 242  
 DB 2 GGAGGGP 8  
 |||||

RESULT 49  
 AAG97623  
 ID AAG97623 standard; Peptide; 10 AA.  
 XX  
 AC AAG97623;  
 XX  
 DT 18-SEP-2001 (first entry)  
 XX  
 DE Human complementary peptide, SEQ ID NO: 3818.



```
Db          |||||
            2 GGGAGGG 8

RESULT 52
AAG97638
ID AAG97638 standard; Peptide; 10 AA.
XX
AC AAG97638;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 3833.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
WPI; 2001-408419/43.
XX
A set of peptide ligands consisting of specific complementary peptides
to proteins encoded by genes of the human genome, useful in an assay
for screening and identifying of one or more novel peptides which are
drug candidates or pro-drugs -
XX
PS Example 6; Page 593; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands
generated from the human genome. The complementary peptides
interact with their relevant target proteins encoded in the human
genome. They can be used as reagents in drug discovery and as lead
ligands to facilitate drug design and development. The present
sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;
XX
Query Match 1.9%; Score 7; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241
Db          |||||
            3 GGGAGGG 9

RESULT 53
AAG83526
ID AAG83526 standard; Peptide; 10 AA.
XX
AC AAG83526;
XX
DT 11-SEP-2001 (first entry)
XX
DE Arabidopsis thaliana peptide ligand #166.
XX
KW Plant; peptide pesticide; peptide herbicide; agricultural research.
XX
OS Arabidopsis thaliana.
XX
PN WO200142279-A2.
XX
Query Match 1.9%; Score 7; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241
Db          |||||
            3 GGGAGGG 9

RESULT 54
AAG84044
ID AAG84044 standard; Peptide; 10 AA.
XX
AC AAG84044;
XX
DT 11-SEP-2001 (first entry)
XX
DE Arabidopsis thaliana peptide ligand #684.
XX
KW Plant; peptide pesticide; peptide herbicide; agricultural research.
XX
OS Arabidopsis thaliana.
XX
PN WO200142279-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04781.
XX
PR 13-DEC-1999; 99GB-0029469.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
WPI; 2001-381629/40.
XX
A set of peptide ligands for agricultural research and development of
therapeutic agents comprise specific complementary peptides to proteins
encoded by genes of plant genomes -
XX
PS Example 4; Page 126; 201pp; English.
XX
CC The present invention relates to a set of peptide ligands consisting of
specific complementary peptides to proteins encoded by genes of plant
```

```
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04781.
XX
PR 13-DEC-1999; 99GB-0029469.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
WPI; 2001-381629/40.
XX
A set of peptide ligands for agricultural research and development of
therapeutic agents comprise specific complementary peptides to proteins
encoded by genes of plant genomes -
XX
PS Example 4; Page 54; 201pp; English.
XX
CC The present invention relates to a set of peptide ligands consisting of
specific complementary peptides to proteins encoded by genes of plant
genomes. The present sequence is one such peptide from Arabidopsis
thaliana. The peptides of the present invention are useful in an assay to
identify a peptide, especially a peptide pesticide or herbicide. The
peptides are also useful for tools for agricultural research and
development.
XX
SQ Sequence 10 AA;
XX
Query Match 1.9%; Score 7; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241
Db          |||||
            3 GGGAGGG 9

RESULT 54
AAG84044
ID AAG84044 standard; Peptide; 10 AA.
XX
AC AAG84044;
XX
DT 11-SEP-2001 (first entry)
XX
DE Arabidopsis thaliana peptide ligand #684.
XX
KW Plant; peptide pesticide; peptide herbicide; agricultural research.
XX
OS Arabidopsis thaliana.
XX
PN WO200142279-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04781.
XX
PR 13-DEC-1999; 99GB-0029469.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
WPI; 2001-381629/40.
XX
A set of peptide ligands for agricultural research and development of
therapeutic agents comprise specific complementary peptides to proteins
encoded by genes of plant genomes -
XX
PS Example 4; Page 126; 201pp; English.
XX
CC The present invention relates to a set of peptide ligands consisting of
specific complementary peptides to proteins encoded by genes of plant
```

CC genomes. The present sequence is one such peptide from Arabidopsis  
 CC thaliana. The peptides of the present invention are useful in an assay to  
 CC identify a peptide, especially a peptide pesticide or herbicide. The  
 CC peptides are also useful for tools for agricultural research and  
 CC development.

XX SQ Sequence 10 AA;

Query Match 1.9%; Score 7; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241  
 |||||  
 Db 1 GGGAGGG 7

RESULT 55

AAR92855  
 ID AAR92855 standard; peptide; 12 AA.

XX AC AAR92855;

XX DT 03-OCT-1996 (first entry)

XX DE Cell adhesion modulatory peptide AP9.

XX KW Intercellular adhesion; stimulation; inhibition; skin graft;  
 KW synthetic blood vessel; coating; endothelial cell; epidermal cell;  
 KW chemotactic attractor; wound healing; organ transplantation;  
 KW thrombosis; arteriosclerosis; cancer metastases.

XX OS Synthetic.

XX PN DE4430601-A1.

XX FD 29-FEB-1996.

XX PF 22-AUG-1994; 94DE-4430601.

XX PR 22-AUG-1994; 94DE-4430601.

XX FA (BEIE ) BEIERSDORF AG.

XX PI Doerschner A, Eichner W, Kock K, Mielke H;

XX DR WPI; 1996-130242/14.

XX PT Peptide(s) that stimulate or inhibit cell to cell adhesion - used  
 PT e.g. to coat synthetic blood vessels with endothelial cells, to  
 PT prepare, or increase growth of skin grafts, to prevent thrombosis  
 PT etc.

XX PS Claim 12; Page 14; 18pp; German.

XX CC The present peptide is a specifically claimed example of a peptide  
 CC which contains the highly generic sequence AA5-AA4-AA3-AA2-AA1-(AAAX)n  
 CC where AA5 is Glu, Ser, Asp or Asn; AA4 is Leu or Ser; AA3 is Leu, Ile,  
 CC phe or Gly; AA2 is Asp, Leu, Asn or Ser; AA1 is Gly, Pro or Asp; AAX  
 CC is any amino acid and n = 0 or 1. When two or more such peptides are  
 CC attached to a carrier, the product can be used for stimulating  
 CC adhesion of eukaryotic cells in vitro. Particular applications include  
 CC coating synthetic blood vessels with endothelial cells, preparing skin  
 CC grafts using epithelial cells or stimulating wound healing. When a  
 CC single peptide is used it may inhibit intercellular adhesion, making  
 CC it useful for preventing thrombosis or arteriosclerosis or to suppress  
 CC cancer metastases. The peptides can also be used as chemotactic  
 CC attractors and for detecting/quantifying cell-cell adhesion in vitro.

XX SQ Sequence 12 AA;

Query Match 1.9%; Score 7; DB 17; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 16;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 235 GGGAGGG 241  
 |||||  
 Db 1 GGGAGGG 7

RESULT 56

AAR24435  
 ID AAR24435 standard; peptide; 14 AA.

XX AC AAR24435;

XX DT 21-NOV-1992 (first entry)

XX DB Sequence of synthetic Epstein-Barr virus (EBV)-encoded nuclear  
 XX antigen (EBNA) designated D3.

XX KW Herpes virus; Epstein-Barr virus; infectious mononucleosis;

XX KM Burkett's lymphoma; nasopharyngeal carcinoma; rheumatoid arthritis.

XX OS Synthetic.

XX PN US5116725-A.

XX PD 26-MAY-1992.

XX PF 04-NOV-1987; 87US-0117241.

XX PR 08-AUG-1984; 84US-0638726.

XX PR 24-MAR-1987; 87US-0029860.

XX PR 04-NOV-1987; 87US-0117241.

XX PA (SCRI ) SCRIPPS CLINIC & RES FOUND.

XX PI Carson DA, Geltosky JE, Houghten R, Rhodes G, Smith RS;

XX PI Vaughan JH;

XX DR WPI; 1992-199513/24.

XX PT Determining state of Epstein-Barr virus infection - by  
 PT determining ratio of IgG to IgM antibodies to synthetic peptides  
 PT in patient

XX PS Claim Iaviii; Column 54; 40pp; English.

XX CC EBNA-1 is the earliest EBV-associated antigen that can be detected  
 CC after infection. EBNA has been detected in the nucleus of latently-  
 CC infected growth-transformed B lymphocytes. EBNA has also been  
 CC detected in the nuclei of African Burkitt tumor lymphoblasts and  
 CC anaplastic nasopharyngeal carcinoma cells. The size of the EBNA-1  
 CC protein is correlated with the variation of the length of the IR-3  
 CC region of EBV-DNA. The IR-3 region encodes a repeating glycine-  
 CC alanine sequence that has been characterised to be the major epitope  
 CC of the EBNA-1 protein. The peptides of the invention are synthetic,  
 CC random copolymeric polypeptides capable of inducing the production of  
 CC antibodies that immunoreact with EBNA and of immunoreacting with  
 CC human antibodies induced by EBNA. The sequences were chosen to  
 CC represent different areas from within and just outside the proposed  
 CC IR-3 region of the EBNA.

XX SQ Sequence 14 AA;

Query Match 1.9%; Score 7; DB 13; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 240  
 |||||  
 Db 6 AGGAGGG 12

RESULT 57



PI	Jakovovits A, Challita-Bid PM, Faris M, Ge W, Hubert RS;	PI	Jakovovits A, Challita-Bid PM, Faris M, Ge W, Hubert RS;
PI	Morrison K, Morrison RK, Raitano AB;	PI	Morrison K, Morrison RK, Raitano AB;
XX		XX	
DR	WPI; 2003-075555/07.	DR	WPI; 2003-075555/07.
XX		XX	
XX	New composition comprising a substance that modulates the structure of	XX	New composition comprising a substance that modulates the structure of
PT	proteins and polynucleotides, useful for therapeutic, prognostic and	PT	proteins and polynucleotides, useful for therapeutic, prognostic and
PT	diagnostic reagents for eliciting cellular or humoral immune response	PT	diagnostic reagents for eliciting cellular or humoral immune response
PT	in cancer patients	PT	in cancer patients
XX		XX	
PS	Claim 13; Page 617; 1021pp; English.	PS	Claim 13; Page 617; 1021pp; English.
XX		XX	
CC	The present invention relates to novel human cancer-related genes and	CC	The present invention relates to novel human cancer-related genes and
CC	proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and	CC	proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and
CC	proteins are useful for eliciting a humoral or cellular immune response.	CC	proteins are useful for eliciting a humoral or cellular immune response.
CC	The genes are useful as probes and primers for the amplification and/or	CC	The genes are useful as probes and primers for the amplification and/or
CC	detection of genes, mRNAs or their fragments, as reagents for the	CC	detection of genes, mRNAs or their fragments, as reagents for the
CC	diagnosis and/or prognosis of cancer, as coding sequences capable of	CC	diagnosis and/or prognosis of cancer, as coding sequences capable of
CC	directing the expression of the protein, as tools for modulating or	CC	directing the expression of the protein, as tools for modulating or
CC	inhibiting the expression of genes and/or translation of transcripts, and	CC	inhibiting the expression of genes and/or translation of transcripts, and
CC	as therapeutic agents. The proteins and peptides are useful as	CC	as therapeutic agents. The proteins and peptides are useful as
CC	therapeutic, prognostic and diagnostic reagents for cancer. The present	CC	therapeutic, prognostic and diagnostic reagents for cancer. The present
CC	sequence is a human leukocyte antigen (HLA) peptide, used in an example	CC	sequence is a human leukocyte antigen (HLA) peptide, used in an example
CC	from the invention.	CC	from the invention.
XX		XX	
XX	Sequence 15 AA;	XX	Sequence 15 AA;
Query Match	1.9%; Score 7; DB 24; Length 15;	Query Match	1.9%; Score 7; DB 24; Length 15;
Best Local Similarity	100.0%; Pred. No. 19;	Best Local Similarity	100.0%; Pred. No. 19;
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	235 GGGAGGG 241	Qy	235 GGGAGGG 241
Db	9 GGGAGGG 15	Db	9 GGGAGGG 15
RESULT 60		RESULT 60	
ABR38300		ABR38300	
ID	ABR38300 standard; Peptide; 15 AA.	ID	ABR38300 standard; Peptide; 15 AA.
AC	ABR38300;	AC	ABR38300;
XX		XX	
XX	19-MAY-2003 (first entry)	XX	19-MAY-2003 (first entry)
DT		DT	
DE	Human cancer-related protein 187P3P2 HLA peptide #1431.	DE	Human cancer-related protein 187P3P2 HLA peptide #1431.
XX		XX	
KW	Human; cytostatic; vaccine; cancer; immune response; HLA;	KW	Human; cytostatic; vaccine; cancer; immune response; HLA;
KW	human leukocyte antigen.	KW	human leukocyte antigen.
XX		XX	
OS	Homo sapiens.	OS	Homo sapiens.
XX		XX	
PN	WO200283921-A2.	PN	WO200283921-A2.
XX		XX	
PD	24-OCT-2002.	PD	24-OCT-2002.
XX		XX	
PF	10-APR-2002; 2002WO-US11654.	PF	10-APR-2002; 2002WO-US11654.
XX		XX	
PR	10-APR-2001; 2001US-282739P.	PR	10-APR-2001; 2001US-282739P.
PR	10-APR-2001; 2001US-283112P.	PR	10-APR-2001; 2001US-283112P.
PR	25-APR-2001; 2001US-286630P.	PR	25-APR-2001; 2001US-286630P.
XX		XX	
XX	(AGEN-) AGENSYS INC.	XX	(AGEN-) AGENSYS INC.
PA		PA	
XX	Jakovovits A, Challita-Bid PM, Faris M, Ge W, Hubert RS;	XX	Jakovovits A, Challita-Bid PM, Faris M, Ge W, Hubert RS;
PI	Morrison K, Morrison RK, Raitano AB;	PI	Morrison K, Morrison RK, Raitano AB;
XX		XX	
XX	WPI; 2003-075555/07.	XX	WPI; 2003-075555/07.
XX		XX	
XX	New composition comprising a substance that modulates the structure of	XX	New composition comprising a substance that modulates the structure of
PT	proteins and polynucleotides, useful for therapeutic, prognostic and	PT	proteins and polynucleotides, useful for therapeutic, prognostic and
PT	diagnostic reagents for eliciting cellular or humoral immune response	PT	diagnostic reagents for eliciting cellular or humoral immune response
PT	in cancer patients	PT	in cancer patients
XX		XX	
PS	Claim 13; Page 617; 1021pp; English.	PS	Claim 13; Page 617; 1021pp; English.
XX		XX	
CC	The present invention relates to novel human cancer-related genes and	CC	The present invention relates to novel human cancer-related genes and
CC	proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and	CC	proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and
CC	proteins are useful for eliciting a humoral or cellular immune response.	CC	proteins are useful for eliciting a humoral or cellular immune response.
CC	The genes are useful as probes and primers for the amplification and/or	CC	The genes are useful as probes and primers for the amplification and/or
CC	detection of genes, mRNAs or their fragments, as reagents for the	CC	detection of genes, mRNAs or their fragments, as reagents for the
CC	diagnosis and/or prognosis of cancer, as coding sequences capable of	CC	diagnosis and/or prognosis of cancer, as coding sequences capable of
CC	directing the expression of the protein, as tools for modulating or	CC	directing the expression of the protein, as tools for modulating or
CC	inhibiting the expression of genes and/or translation of transcripts, and	CC	inhibiting the expression of genes and/or translation of transcripts, and
CC	as therapeutic agents. The proteins and peptides are useful as	CC	as therapeutic agents. The proteins and peptides are useful as



CC as therapeutic agents. The proteins and peptides are useful as  
 CC therapeutic, prognostic and diagnostic reagents for cancer. The present  
 CC sequence is a human leukocyte antigen (HLA) peptide, used in an example  
 CC from the invention.

XX SQ Sequence 15 AA;

Query Match 1.9%; Score 7; DB 24; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
 |||||  
 Db 9 GGGAGGG 15

RESULT 62

ABR38314  
 ID ABR38314 standard; Peptide; 15 AA.

XX AC ABR38314;

XX XX 19-MAY-2003 (first entry)

XX DE Human cancer-related protein 187P3F2 HLA peptide #1445.

XX XX Human; cytostatic; vaccine; cancer; immune response; HLA;  
 KW human leukocyte antigen.

XX OS Homo sapiens.

XX PN WO200283921-A2.

XX PD 24-OCT-2002.

XX PF 10-APR-2002; 2002WO-US11654.

XX PR 10-APR-2001; 2001US-282739P.

XX PR 10-APR-2001; 2001US-283112P.

XX PR 25-APR-2001; 2001US-286630P.

XX PA (AGEN-) AGENSYS INC.

XX PI Jakobovits A, Challita-Eid PM, Faris M, Ge W, Hubert RS;

XX PI Morrison K, Morrison RK, Raitano AB;

XX DR WPI; 2003-075555/07.

XX PT New composition comprising a substance that modulates the structure of  
 PT proteins and polynucleotides, useful for therapeutic, prognostic and  
 PT diagnostic reagents for eliciting cellular or humoral immune response  
 PT in cancer patients -

XX PS Claim 13; Page 617; 1021pp; English.

XX CC The present invention relates to novel human cancer-related genes and  
 CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and  
 CC proteins are useful for eliciting a humoral or cellular immune response.  
 CC The genes are useful as probes and primers for the amplification and/or  
 CC detection of genes, mRNAs or their fragments, as reagents for the  
 CC diagnosis and/or prognosis of cancer, as coding sequences capable of  
 CC directing the expression of the protein, as tools for modulating or  
 CC inhibiting the expression of genes and/or translation of transcripts, and  
 CC as therapeutic agents. The proteins and peptides are useful as  
 CC therapeutic, prognostic and diagnostic reagents for cancer. The present  
 CC sequence is a human leukocyte antigen (HLA) peptide, used in an example  
 CC from the invention.

XX SQ Sequence 15 AA;

Query Match 1.9%; Score 7; DB 24; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
 |||||  
 Db 5 GGGAGGG 11

RESULT 63

ABR38327  
 ID ABR38327 standard; Peptide; 15 AA.

XX AC ABR38327;

XX DT 19-MAY-2003 (first entry)

XX DE Human cancer-related protein 187P3F2 HLA peptide #1458.

XX KW Human; cytostatic; vaccine; cancer; immune response; HLA;  
 KW human leukocyte antigen.

XX OS Homo sapiens.

XX PN WO200283921-A2.

XX PD 24-OCT-2002.

XX PF 10-APR-2002; 2002WO-US11654.

XX PR 10-APR-2001; 2001US-282739P.

XX PR 10-APR-2001; 2001US-283112P.

XX PR 25-APR-2001; 2001US-286630P.

XX PA (AGEN-) AGENSYS INC.

XX PI Jakobovits A, Challita-Eid PM, Faris M, Ge W, Hubert RS;

XX PI Morrison K, Morrison RK, Raitano AB;

XX DR WPI; 2003-075555/07.

XX PT New composition comprising a substance that modulates the structure of  
 PT proteins and polynucleotides, useful for therapeutic, prognostic and  
 PT diagnostic reagents for eliciting cellular or humoral immune response  
 PT in cancer patients -

XX PS Claim 13; Page 617; 1021pp; English.

XX CC The present invention relates to novel human cancer-related genes and  
 CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and  
 CC proteins are useful for eliciting a humoral or cellular immune response.  
 CC The genes are useful as probes and primers for the amplification and/or  
 CC detection of genes, mRNAs or their fragments, as reagents for the  
 CC diagnosis and/or prognosis of cancer, as coding sequences capable of  
 CC directing the expression of the protein, as tools for modulating or  
 CC inhibiting the expression of genes and/or translation of transcripts, and  
 CC as therapeutic agents. The proteins and peptides are useful as  
 CC therapeutic, prognostic and diagnostic reagents for cancer. The present  
 CC sequence is a human leukocyte antigen (HLA) peptide, used in an example  
 CC from the invention.

XX SQ Sequence 15 AA;

Query Match 1.9%; Score 7; DB 24; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
 |||||  
 Db 3 GGGAGGG 9

RESULT 64

ABR38328  
 ID ABR38328 standard; Peptide; 15 AA.

XX

AC ABR38328;  
 XX 19-MAY-2003 (first entry)  
 DT Human cancer-related protein 187P3F2 HLA peptide #1459.  
 DE Human; cytostatic; vaccine; cancer; immune response; HLA;  
 DE human leukocyte antigen.  
 KW Homo sapiens.  
 XX WO200283921-A2.  
 XX 24-OCT-2002.  
 XX 10-APR-2002; 2002WO-US11654.  
 XX 10-APR-2001; 2001US-282739P.  
 PR 10-APR-2001; 2001US-283112P.  
 PR 25-APR-2001; 2001US-286630P.  
 XX (AGEN-) AGENSYS INC.  
 XX Jakobovits A, Challita-Eid PM, Faris M, Ge W, Hubert RS;  
 PI Morrison K, Morrison RK, Raitano AB;  
 PI WPI; 2003-075555/07.  
 DR New composition comprising a substance that modulates the structure of  
 PT proteins and polynucleotides, useful for therapeutic, prognostic and  
 PT diagnostic reagents for eliciting cellular or humoral immune response  
 PT in cancer patients  
 XX Claim 13; Page 617; 1021pp; English.  
 XX The present invention relates to novel human cancer-related genes and  
 CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and  
 CC proteins are useful for eliciting a humoral or cellular immune response.  
 CC The genes are useful as probes and primers for the amplification and/or  
 CC detection of genes, mRNAs or their fragments, as reagents for the  
 CC diagnosis and/or prognosis of cancer, as coding sequences capable of  
 CC directing the expression of the protein, as tools for modulating or  
 CC inhibiting the expression of genes and/or translation of transcripts, and  
 CC as therapeutic agents. The proteins and peptides are useful as  
 CC agents for eliciting the expression of genes and/or translation of transcripts, and  
 CC as therapeutic, prognostic and diagnostic reagents for cancer. The present  
 CC sequence is a human leukocyte antigen (HLA) peptide, used in an example  
 CC from the invention.  
 XX Query Match 1.9%; Score 7; DB 24; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 235 GGGAGGG 241  
 Db |||||  
 2 GGGAGGG 8  
 Sequence 15 AA;  
 RESULT 65  
 ABR38464  
 ID ABR38464 standard; Peptide; 15 AA.  
 XX ABR38464;  
 AC ABR38464;  
 XX 19-MAY-2003 (first entry)  
 DT Human cancer-related protein 187P3F2 HLA peptide #1595.  
 DE Human; cytostatic; vaccine; cancer; immune response; HLA;  
 DE human leukocyte antigen.  
 KW Homo sapiens.  
 XX

XX WO200283921-A2.  
 XX 24-OCT-2002.  
 XX 10-APR-2002; 2002WO-US11654.  
 XX 10-APR-2001; 2001US-282739P.  
 PR 10-APR-2001; 2001US-283112P.  
 PR 25-APR-2001; 2001US-286630P.  
 XX (AGEN-) AGENSYS INC.  
 XX Jakobovits A, Challita-Eid PM, Faris M, Ge W, Hubert RS;  
 PI Morrison K, Morrison RK, Raitano AB;  
 PI WPI; 2003-075555/07.  
 DR New composition comprising a substance that modulates the structure of  
 PT proteins and polynucleotides, useful for therapeutic, prognostic and  
 PT diagnostic reagents for eliciting cellular or humoral immune response  
 PT in cancer patients  
 XX Claim 13; Page 619; 1021pp; English.  
 XX The present invention relates to novel human cancer-related genes and  
 CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and  
 CC proteins are useful for eliciting a humoral or cellular immune response.  
 CC The genes are useful as probes and primers for the amplification and/or  
 CC detection of genes, mRNAs or their fragments, as reagents for the  
 CC diagnosis and/or prognosis of cancer, as coding sequences capable of  
 CC directing the expression of the protein, as tools for modulating or  
 CC inhibiting the expression of genes and/or translation of transcripts, and  
 CC as therapeutic agents. The proteins and peptides are useful as  
 CC agents for eliciting the expression of genes and/or translation of transcripts, and  
 CC as therapeutic, prognostic and diagnostic reagents for cancer. The present  
 CC sequence is a human leukocyte antigen (HLA) peptide, used in an example  
 CC from the invention.  
 XX Query Match 1.9%; Score 7; DB 24; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 235 GGGAGGG 241  
 Db |||||  
 2 GGGAGGG 8  
 Sequence 15 AA;  
 RESULT 66  
 AAR24433  
 ID AAR24433 standard; peptide; 16 AA.  
 XX AAR24433;  
 AC AAR24433;  
 XX 21-NOV-1992 (first entry)  
 DT Sequence of synthetic Epstein-Barr virus (EBV)-encoded nuclear  
 DE antigen (EBNA) designated F15.  
 DE Herpes virus; Epstein-Barr virus; infectious mononucleosis;  
 KW Burkett's lymphoma; nasopharyngeal carcinoma; rheumatoid arthritis.  
 XX Synthetic.  
 XX US5116725-A.  
 XX 26-MAY-1992.  
 PD 04-NOV-1987; 87US-0117241.  
 PF 08-AUG-1984; 84US-0638726.  
 PR 24-MAR-1987; 87US-0029860.  
 PR



CC infected growth-transformed B lymphocytes. EBNA has also been  
 CC detected in the nuclei of African Burkitt tumor lymphoblasts and  
 CC anaplastic nasopharyngeal carcinoma cells. The size of the EBNA-1  
 CC protein is correlated with the variation of the length of the IR-3  
 CC region of EBV-DNA. The IR-3 region encodes a repeating glycine-  
 CC alanine sequence that has been characterised to be the major epitope  
 CC of the EBNA-1 protein. The peptides of the invention are synthetic,  
 CC random copolymeric polypeptides capable of inducing the production of  
 CC antibodies that immunoreact with EBNA and of immunoreacting with  
 CC human antibodies induced by EBNA. The sequences were chosen to  
 CC represent different areas from within and just outside the proposed  
 CC IR-3 region of the EBNA.

XX Sequence 17 AA;

Query Match 1.9%; Score 7; DB 13; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGG 240  
 |||||  
 Db 9 AGGGAGG 15

#### RESULT 69

AAAR25056  
 ID AAR25056 standard; Protein; 17 AA.

XX AC AAR25056;

DT 25-MAR-2003 (updated)

DT 09-DEC-1992 (first entry)

XX Synthetic random copolymer peptide #1.

DE EBNA; antibody; IM; EBV; cytomegalovirus.

XX Synthetic.

XX US5122448-A.

PD 16-JUN-1992.

XX 11-JAN-1990; 90US-0463505.

XX 08-AUG-1984; 84US-0638726.

PR 24-MAR-1987; 87US-0029860.

PR 04-NOV-1987; 87US-0117241.

PR 11-JAN-1990; 90US-0463505.

XX (SCRI ) SCRIPPS CLINIC & RES FOUND.

XX Carson DA, Houghten R, Rhodes G, Vaughan JH;

DR WPI; 1992-226077/27.

XX Assay for anti-Epstein Barr virus nuclear antigen antibodies -  
 PT using random copolymer polypeptide contg. specified aminoacid  
 PT residues, for diagnosis of infectious mononucleosis and  
 PT nasopharyngeal carcinoma

PS Claim 1; Page 54; 41pp; English.

XX The sequences given in AAR25050-7 are random copolymer peptides which  
 CC can be used in diagnostic methods and systems relating to Epstein-  
 CC Barr virus nuclear antigen (EBNA). These peptides are capable of  
 CC inducing the production of antibodies which immunoreact with EBNA,  
 CC and of immunoreacting with human antibodies induced by EBNA. The  
 CC peptides can be fixed to a solid matrix, giving a solid  
 CC immunoreactant in a diagnostic kit. The assay developed using  
 CC these peptides has been found to be clinically reliable in  
 CC detecting infectious mononucleosis (IM) caused by Epstein-Barr virus  
 CC (EBV) as well as IM induced by cytomegalovirus, and also in

CC detecting nasopharyngeal carcinoma, another disease in which EBV  
 CC has been implicated.  
 CC (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 17 AA;

Query Match 1.9%; Score 7; DB 13; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGG 240  
 |||||  
 Db 9 AGGGAGG 15

#### RESULT 70

AAW79101

ID AAW79101 standard; peptide; 17 AA.

XX AC AAW79101;

DT 18-NOV-1998 (first entry)

XX Stabilising sequence of the invention.

XX Fusion protein; stabilising polypeptide; proteolytic degradation;  
 KW resistance; half-life; autoimmune disease; inflammation; nitro drug;  
 KW IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;  
 KW nitroreductase protein; enzyme therapy; prodrug therapy; protease;  
 KW cancer; pathological condition.

XX Epstein-barr virus.

XX WO9822577-A1.

XX 28-MAY-1998.

XX 17-NOV-1997; 97WO-IB01508.

XX 25-JUN-1997; 97US-0048945.

PR 15-NOV-1996; 96US-0030986.

XX (MASU/) MASUCCI M G.

XX Masucci MG;

XX WPI; 1998-312463/27.

DR N-PSDB; AAV55829.

XX New fusion proteins resistant to proteolytic degradation -  
 PT comprising a core protein with a stabilising polypeptide comprising  
 PT a peptide sequence containing glycine repeats

XX Disclosure; Fig 2; 120pp; English.

XX This represents a stabilising sequence of the invention. The invention  
 CC provides a method for increasing the resistance of a core protein to  
 CC proteolytic degradation that comprises linking or inserting onto or into  
 CC the core protein a stabilising polypeptide of formula [(Glya)(Glyb)y  
 CC (Glyc)Z]n where Glya, Glyb, Glyc are 1-6 sequential Gly residues and  
 CC x, y, z are Ala, Ser, Val, Ile, Leu, Met, Phe, Pro or Thr and n can be  
 CC anything between 1-66. X, Y and Z need not be identical from n repeat to  
 CC n repeat. Alternatively a nucleic acid encoding the stabilising  
 CC polypeptide can be linked onto or inserted into a nucleic acid encoding  
 CC a core protein. The fusion proteins of the invention are more resistant  
 CC to degradation by proteases and, thus, have a longer half-life than the  
 CC unfused core protein. The products can be used for treating autoimmune  
 CC diseases, cancer and inflammation. In particular, the core protein may  
 CC be an IkappaB regulator protein for the treatment of inflammatory bowel  
 CC disease, or a nitroreductase protein which can activate nitro drugs in  
 CC enzyme/prodrug therapy to treat cancer or other pathological conditions.  
 CC The fusion proteins can also be used in diagnostic methods such as in  
 CC vivo imaging.

```

XX SQ Sequence 17 AA;
Query Match 1.9%; Score 7; DB 19; Length 17;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGG 240
DB 11 AGGGAGG 17
|||||

RESULT 71
ABE76028
ID ABB76028 standard; Peptide; 19 AA.
XX AC ABB76028;
XX DT 12-JUL-2002 (first entry)
XX DE Streptococcus pneumoniae mucin-binding protein peptide fragment.
XX KW Mucin-binding protein; vaccine; diagnosis; gene therapy;
XX KW otitis media; rhinosinusitis; bacteraemia; meningitis; pneumonia;
XX KW lower respiratory tract infection; antimicrobial; antibiotic;
XX KW auditory; immunostimulant; vaccine.
XX OS Streptococcus pneumoniae.
XX PN WO200228351-A2.
XX PD 11-APR-2002.
XX PF 04-OCT-2001; 2001WO-US311269.
XX PR 04-OCT-2000; 2000US-237888P.
XX PR 07-FEB-2001; 2001US-267104P.
XX PA (AMHP ) AMERICAN HOME PROD CORP.
XX PA (UUNY ) UNIV NEW YORK STATE RES FOUND.
XX PI Green BA, Masi AW, Reddy MS;
XX WPI; 2002-383318/41.
XX DR Mucin binding proteins, useful in the induction of an immune response
XX PT to, and in the diagnosis of, pneumococcal infections -
XX PS Example; Page 69; 71pp; English.
XX CC The present sequence is an N-terminal peptide fragment of a 14 kDa
XX CC mucin-binding protein (see ABB76026) from the outer membrane
XX CC fraction of Streptococcus pneumoniae strain Cp1200. Elucidation
XX CC of this N-terminal sequence allowed the identification of the
XX CC corresponding open reading frame (see ABL54193) in the S.
XX CC pneumoniae genome. The invention provides amino acid and nucleic
XX CC acid sequences for isolated 12 kDa and 14 kDa mucin-binding proteins,
XX CC as well as expression vectors, transfected host cells, methods for
XX CC producing recombinant mucin-binding proteins, compositions comprising
XX CC the proteins, and antibodies to the proteins. Claimed compositions
XX CC comprising either a pneumococcal mucin-binding protein or an
XX CC expression vector encoding such a mucin-binding protein are used to
XX CC elicit protective immunity from a disease caused by S. pneumoniae,
XX CC especially otitis media, rhinosinusitis, bacteraemia, meningitis,
XX CC pneumonia or lower respiratory tract infection (all claimed).
XX SQ Sequence 19 AA;
Query Match 1.9%; Score 7; DB 23; Length 19;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 VKKNIEE 32

```

```

DB 9 VKKNIEE 15
|||||

RESULT 72
AAP60475
ID AAP60475 standard; protein; 20 AA.
XX AC AAP60475;
XX DT 25-MAR-2003 (updated)
XX DT 28-JUL-1991 (first entry)
XX DE Epstein-Barr virus nuclear antigen (EBNA) pentapeptide.
XX KW Epstein-Barr virus; nuclear antigen; passive immunization; diagnosis.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Region 2..6
XX FT /label= EBNA determinant
XX PN WO8601210-A.
XX PD 27-FEB-1986.
XX PF 02-AUG-1985; 85WO-US01484.
XX PR 08-AUG-1984; 84US-0638726.
XX PA (SCRI ) SCRIPPS CLINIC & RES FOUND.
XX PI Vaughan JH, Carbon DA, Rhodes G, Houghten R;
XX WPI; 1986-068962/10.
XX PT New synthetic copolymer polypeptide(s) - useful for preventing
XX PT and diagnosing diseases involving Epstein-Barr virus.
XX PS Claim 4(iii); Page 73; 84pp; English.
XX CC The peptide includes the EBNA pentapeptide which is an
XX CC immunogen for the production of an antibody in a host animal.
XX CC Epstein-Barr virus- and EBNA-associated diseases may be
XX CC diagnosed and prevented
XX CC See also AAP60473-P60483
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 20 AA;
Query Match 1.9%; Score 7; DB 7; Length 20;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGG 240
DB 3 AGGGAGG 9
|||||

RESULT 73
AAR03360
ID AAR03360 standard; protein; 20 AA.
XX AC AAR03360;
XX DT 25-MAR-2003 (updated)
XX DT 31-JUL-1990 (first entry)
XX DE Antigenic fragment of CMV-encoded polypeptide.
XX KW CMV; Epstein-Barr virus; EBV; infectious mononucleosis; IN;
XX KW nasopharyngeal carcinoma; NPC.

```

```

XX OS Cytomegalovirus.
XX PN WO9001495-A.
XX PD 22-FEB-1990.
XX PF 03-AUG-1989; 89WO-US03350.
XX PR 08-AUG-1988; 88US-0230250.
XX PA (SCRI ) SCRIPPS CLINIC & RE.
XX PI Rhodes G, Smith RS;
XX DR WPI; 1990-083477/11.
XX PT Synthetic peptide(s) corresp. to cytomegalovirus -
PT used to detect Epstein-Barr virus nuclear antigens and for
PT prodn. of antibodies.
XX PS Claim 3; Page 108; 135pp; English.
XX CC Peptide derivatives and antigenically related variants may be used as an
CC assay or to generate antibodies for passive immunisation against Epstein-
CC Barr virus encoded nuclear antigens, infectious mononucleosis and
CC nasopharyngeal carcinoma.
CC (Updated on 25-MAR-2003 to correct PF field.)
CC (Updated on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 20 AA;

Query Match 1.9%; Score 7; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241
Db |||||
6 GGGAGGG 12

RESULT 74
AAR24430
ID AAR24430 standard; peptide; 20 AA.
XX AC AAR24430;
XX DT 21-NOV-1992 (first entry)
XX DE Sequence of synthetic Epstein-Barr virus (EBV)-encoded nuclear
DE antigen (EBNA) designated P62(B).
XX KW Herpes virus; Epstein-Barr virus; infectious mononucleosis;
KW Burkett's lymphoma; nasopharyngeal carcinoma; rheumatoid arthritis.
XX OS Synthetic.
XX PN US5116725-A.
XX PD 26-MAY-1992.
XX PF 04-NOV-1987; 87US-0117241.
XX PR 08-AUG-1984; 84US-0638726.
XX PR 24-MAR-1987; 87US-0029860.
XX PR 04-NOV-1987; 87US-0117241.
XX PA (SCRI ) SCRIPPS CLINIC & RES FOUND.
XX PI Carson DA, Geltosky JE, Houghten R, Rhodes G, Smith RS;
XX PI Vaughan JH;
XX DR WPI; 1992-199513/24.

```

```

XX PT Determining state of Epstein-Barr virus infection - by
PT determining ratio of IgG to IgM antibodies to synthetic peptides
PT in patient
XX PS Claim 1a1ii; Column 54; 40pp; English.
XX CC EBNA-1 is the earliest EBV-associated antigen that can be detected
CC after infection. EBNA has been detected in the nucleus of latently-
CC infected growth-transformed B lymphocytes. EBNA has also been
CC detected in the nuclei of African Burkitt tumor lymphoblasts and
CC anaplastic nasopharyngeal carcinoma cells. The size of the EBNA-1
CC protein is correlated with the variation of the length of the IR-3
CC region of EBV-DNA. The IR-3 region encodes a repeating glycine-
CC alanine sequence that has been characterised to be the major epitope
CC of the EBNA-1 protein. The peptides of the invention are synthetic,
CC random copolymeric polypeptides capable of inducing the production of
CC antibodies that immunoreact with EBNA and of immunoreacting with
CC human antibodies induced by EBNA. The sequences were chosen to
CC represent different areas from within and just outside the proposed
CC IR-3 region of the EBNA.
XX SQ Sequence 20 AA;

Query Match 1.9%; Score 7; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGG 240
Db |||||
3 AGGGAGG 9

RESULT 75
AAR25052
ID AAR25052 standard; Protein; 20 AA.
XX AC AAR25052;
XX DT 25-MAR-2003 (updated)
DT 09-DEC-1992 (first entry)
XX DE Synthetic random copolymer peptide P62.
XX KW EBNA; antibody; IM; EBV; cytomegalovirus.
XX OS Synthetic.
XX PN US5122448-A.
XX PD 16-JUN-1992.
XX PF 11-JAN-1990; 90US-0463505.
XX PR 08-AUG-1984; 84US-0638726.
XX PR 24-MAR-1987; 87US-0029860.
XX PR 04-NOV-1987; 87US-0117241.
XX PR 11-JAN-1990; 90US-0463505.
XX PA (SCRI ) SCRIPPS CLINIC & RES FOUND.
XX PI Carson DA, Houghten R, Rhodes G, Vaughan JH;
XX DR WPI; 1992-226077/27.
XX PT Assay for anti-Epstein Barr virus nuclear antigen antibodies -
PT using random copolymer polypeptide contg. specified aminoacid
PT residues, for diagnosis of infectious mononucleosis and
PT nasopharyngeal carcinoma
XX PS Claim 1; Page 54; 41pp; English.
XX CC The sequences given in AAR25050-7 are random copolymer peptides which

```

CC can be used in diagnostic methods and systems relating to Epstein-  
CC Barr virus nuclear antigen (EBNA). These peptides are capable of  
CC inducing the production of antibodies which immunoreact with EBNA,  
CC and of immunoreacting with human antibodies induced by EBNA. The  
CC peptides can be fixed to a solid matrix, giving a solid  
CC immunoreactant in a diagnostic kit. The assay developed using  
CC these peptides has been found to be clinically reliable in  
CC detecting infectious mononucleosis (IM) caused by Epstein-Barr virus  
CC (EBV) as well as IM induced by cytomegalovirus, and also in  
CC detecting nasopharyngeal carcinoma, another disease in which EBV  
CC has been implicated.  
CC (Updated on 25-MAR-2003 to correct PF field.)  
XX  
SQ Sequence 20 AA;  
Query Match 1.9%; Score 7; DB 13; Length 20;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 234 AGGAGG 240  
| | | | |  
Db 3 AGGAGG 9  
Search completed: December 17, 2003, 06:33:21  
Job time : 37.7529 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 06:35:27 ; Search time 30.9444 Seconds  
(without alignments)  
2220.991 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 368

Sequence: 1 MKLTVKTLKGTHTFIRVQPN.....CDRNEELAANYLLEHAGEED 368

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 696363 seqs, 186758610 residues

Word size : 0

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	368	100.0	368	9	US-09-805-550-4
2	13	3.5	405	9	US-09-805-550-2
3	8	2.2	10	12	US-09-572-270A-6
4	8	2.2	10	12	US-09-572-270A-8
5	8	2.2	177	12	US-10-029-386-32842
6	8	2.2	306	12	US-10-267-430-13
7	8	2.2	325	12	US-10-267-430-4
8	8	2.2	331	12	US-10-194-885-10
9	8	2.2	332	12	US-10-091-841-9
10	8	2.2	401	15	US-10-295-403-148
11	8	2.2	744	12	US-10-063-685-67
12	8	2.2	744	15	US-10-184-644-255
13	8	2.2	744	15	US-10-184-634-255
14	8	2.2	755	12	US-10-140-472-153
15	8	2.2	755	12	US-10-141-761-153

8	2.2	755	12	US-10-142-885-153	Sequence 153, App
8	2.2	755	12	US-10-158-790-153	Sequence 153, App
8	2.2	755	12	US-10-137-871-153	Sequence 153, App
8	2.2	755	12	US-10-140-805-153	Sequence 153, App
8	2.2	755	12	US-10-140-864-153	Sequence 153, App
8	2.2	755	12	US-10-140-923-153	Sequence 153, App
8	2.2	755	12	US-10-141-756-153	Sequence 153, App
8	2.2	755	12	US-10-141-759-153	Sequence 153, App
8	2.2	755	15	US-10-123-155-153	Sequence 153, App
8	2.2	755	16	US-10-146-731-153	Sequence 153, App
8	2.2	828	12	US-10-174-677-34	Sequence 34, Appl
8	2.2	828	12	US-10-140-472-177	Sequence 177, App
8	2.2	959	12	US-10-141-761-177	Sequence 177, App
8	2.2	959	12	US-10-142-885-177	Sequence 177, App
8	2.2	959	12	US-10-158-790-177	Sequence 177, App
8	2.2	959	12	US-10-137-871-177	Sequence 177, App
8	2.2	959	12	US-10-140-805-177	Sequence 177, App
8	2.2	959	12	US-10-140-864-177	Sequence 177, App
8	2.2	959	12	US-10-140-923-177	Sequence 177, App
8	2.2	959	12	US-10-141-756-177	Sequence 177, App
8	2.2	959	12	US-10-141-759-177	Sequence 177, App
8	2.2	959	15	US-10-123-155-177	Sequence 177, App
8	2.2	959	16	US-10-146-731-177	Sequence 177, App
8	2.2	997	15	US-10-184-644-23	Sequence 23, Appl
8	2.2	997	15	US-10-184-634-23	Sequence 23, Appl
8	2.2	1004	10	US-09-738-626-5676	Sequence 5676, Ap
8	2.2	1016	15	US-10-184-634-99	Sequence 99, Appl
8	2.2	1016	15	US-10-184-634-99	Sequence 99, Appl
8	2.2	1176	12	US-10-063-685-87	Sequence 87, Appl
8	2.2	1176	15	US-10-184-644-293	Sequence 293, App
8	2.2	1176	15	US-10-184-634-293	Sequence 293, App
8	2.2	1200	9	US-09-826-508-3	Sequence 3, Appl
8	2.2	1245	15	US-10-184-644-223	Sequence 223, App
8	2.2	1245	15	US-10-184-634-223	Sequence 223, App
8	2.2	1461	12	US-10-356-153-86	Sequence 86, Appl
8	2.2	1461	15	US-10-142-231-86	Sequence 86, Appl
8	2.2	1503	12	US-10-141-761-363	Sequence 363, App
8	2.2	1503	12	US-10-140-864-363	Sequence 363, App
8	2.2	1503	12	US-10-158-790-363	Sequence 363, App
8	2.2	1503	12	US-10-137-871-363	Sequence 363, App
8	2.2	1503	12	US-10-140-805-363	Sequence 363, App
8	2.2	1503	12	US-10-140-923-363	Sequence 363, App
8	2.2	1503	12	US-10-141-756-363	Sequence 363, App
8	2.2	1503	12	US-10-141-759-363	Sequence 363, App
8	2.2	1503	16	US-10-123-155-363	Sequence 363, App
8	2.2	1524	15	US-10-146-731-363	Sequence 363, App
8	2.2	1524	15	US-10-184-644-421	Sequence 421, App
8	2.2	1564	15	US-10-184-634-245	Sequence 245, App
8	2.2	1564	15	US-10-184-634-245	Sequence 245, App
8	2.2	1572	12	US-10-140-472-373	Sequence 373, App
8	2.2	1572	12	US-10-141-761-373	Sequence 373, App
8	2.2	1572	12	US-10-142-885-373	Sequence 373, App
8	2.2	1572	12	US-10-158-790-373	Sequence 373, App
8	2.2	1572	12	US-10-137-871-373	Sequence 373, App
8	2.2	1572	12	US-10-140-805-373	Sequence 373, App
8	2.2	1572	12	US-10-140-864-373	Sequence 373, App
8	2.2	1572	12	US-10-140-923-373	Sequence 373, App
8	2.2	1572	12	US-10-141-756-373	Sequence 373, App
8	2.2	1572	12	US-10-141-759-373	Sequence 373, App
8	2.2	1572	15	US-10-123-155-373	Sequence 373, App
8	2.2	1572	16	US-10-146-731-373	Sequence 373, App
8	2.2	1638	15	US-10-184-644-49	Sequence 49, Appl
8	2.2	1638	15	US-10-184-634-49	Sequence 49, Appl
8	2.2	1670	12	US-10-140-472-325	Sequence 325, App
8	2.2	1670	12	US-10-141-761-325	Sequence 325, App
8	2.2	1670	12	US-10-142-885-325	Sequence 325, App
8	2.2	1670	12	US-10-158-790-325	Sequence 325, App
8	2.2	1670	12	US-10-137-871-325	Sequence 325, App
8	2.2	1670	12	US-10-140-805-325	Sequence 325, App
8	2.2	1670	12	US-10-140-864-325	Sequence 325, App



89 8 2.2 1670 12 US-10-140-923-325 Sequence 325, App  
90 8 2.2 1670 12 US-10-141-756-325 Sequence 325, App  
91 8 2.2 1670 12 US-10-141-759-325 Sequence 325, App  
92 8 2.2 1670 15 US-10-123-155-325 Sequence 325, App  
93 8 2.2 1670 16 US-10-146-731-325 Sequence 325, App  
94 8 2.2 1671 15 US-10-184-644-505 Sequence 505, App  
95 8 2.2 1671 15 US-10-184-634-505 Sequence 505, App  
96 8 2.2 1679 12 US-10-140-472-375 Sequence 375, App  
97 8 2.2 1679 12 US-10-141-761-375 Sequence 375, App  
98 8 2.2 1679 12 US-10-142-885-375 Sequence 375, App  
99 8 2.2 1679 12 US-10-158-790-375 Sequence 375, App  
100 8 2.2 1679 12 US-10-137-871-375 Sequence 375, App

## ALIGNMENTS

RESULT 1  
US-09-805-550-4  
; Sequence 4, Application US/09805550  
; Patent No. US20020026045A1  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; APPLICANT: Tagliani, Laura  
; FILE REFERENCE: 0964D  
; CURRENT APPLICATION NUMBER: US/09/805,550  
; CURRENT FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 09/413,574  
; PRIOR FILING DATE: 1999-10-06  
; PRIOR APPLICATION NUMBER: 60/109,728  
; PRIOR FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-805-550-4

Query Match 100.0%; Score 368; DB 9; Length 368;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLTVKLTGTHFEIRVQNDTIMAVKKNIEIQKDSYPWQQLIFNGKVLKDESTLE 60  
DB 1 MKLTVKLTGTHFEIRVQNDTIMAVKKNIEIQKDSYPWQQLIFNGKVLKDESTLE 60  
QY 61 ENKVNEDGFLVWLSKGTSGTGTSSQHSNTPATRQAPPLEAQPPVPVPIITTSQ 120  
DB 61 ENKVNEDGFLVWLSKGTSGTGTSSQHSNTPATRQAPPLEAQPPVPVPIITTSQ 120  
QY 121 PGLPQAQNTDHAASNLSSGRVDTIINQLEMGGSDKDKVQALRAAYNNPERAV 180  
DB 121 PGLPQAQNTDHAASNLSSGRVDTIINQLEMGGSDKDKVQALRAAYNNPERAV 180  
QY 181 EYLXSGIPVTAEIAPVIGGGANTTDRAPTGAGLSGIPNTAPLDLFFQASNAGGAGG 240  
DB 181 EYLXSGIPVTAEIAPVIGGGANTTDRAPTGAGLSGIPNTAPLDLFFQASNAGGAGG 240  
QY 241 GPLDLFRNNPQAVREVMYHTNPQILQPLMVELSKONPQILRIENHDEFLLQNEPFE 300  
DB 241 GPLDLFRNNPQAVREVMYHTNPQILQPLMVELSKONPQILRIENHDEFLLQNEPFE 300  
QY 301 GEGDFLDQDEDEMPHAI SVTPQEQAIGRLESWGFDRAVIEAFLACDRNEELAANYL 360  
DB 301 GEGDFLDQDEDEMPHAI SVTPQEQAIGRLESWGFDRAVIEAFLACDRNEELAANYL 360  
QY 361 LEHAGEED 368  
DB 361 LEHAGEED 368

RESULT 2  
US-09-805-550-2  
; Sequence 2, Application US/09805550  
; Patent No. US20020026045A1  
; GENERAL INFORMATION:  
; APPLICANT: Mahajan, Pramod B.  
; APPLICANT: Tagliani, Laura  
; FILE REFERENCE: 0964D  
; CURRENT APPLICATION NUMBER: US/09/805,550  
; CURRENT FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 09/413,574  
; PRIOR FILING DATE: 1999-10-06  
; PRIOR APPLICATION NUMBER: 60/109,728  
; PRIOR FILING DATE: 1998-11-23  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 405  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-805-550-2

Query Match 3.5%; Score 13; DB 9; Length 405;  
Best Local Similarity 100.0%; Pred. No. 0.001;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 RALRAAYNNPERA 179  
DB 182 RALRAAYNNPERA 194

RESULT 3  
US-09-572-270A-6  
; Sequence 6, Application US/09572270A  
; Publication No. US20030148368A1  
; GENERAL INFORMATION:  
; APPLICANT: Proteom Ltd  
; TITLE OF INVENTION: Inter- complementary peptide listing  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/09/572,270A  
; CURRENT FILING DATE: 2000-05-17  
; NUMBER OF SEQ ID NOS: 1144  
; SOFTWARE: ProtPatent version 1.0  
; SEQ ID NO 6  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Arabidopsis Thaliana  
; OTHER INFORMATION: Sequence located in GRP7. at 196-205 and may interact with  
US-09-572-270A-6

Query Match 2.2%; Score 8; DB 12; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
DB 1 AGGGAGGG 8

RESULT 4  
US-09-572-270A-8  
; Sequence 8, Application US/09572270A  
; Publication No. US20030148368A1  
; GENERAL INFORMATION:  
; APPLICANT: Proteom Ltd  
; TITLE OF INVENTION: Inter- complementary peptide listing  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/09/572,270A  
; CURRENT FILING DATE: 2000-05-17  
; NUMBER OF SEQ ID NOS: 1144  
; SOFTWARE: ProtPatent version 1.0  
; SEQ ID NO 8

; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Arabidopsis Thaliana  
; OTHER INFORMATION: Sequence located in GRP7. at 138-147 and may interact with  
US-09-572-270A-8

Query Match 2.2%; Score 8; DB 12; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
|||||  
Db 1 AGGGAGGG 8

## RESULT 5

US-10-029-386-32842  
; Sequence 32842, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEOICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 32842  
; LENGTH: 177  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL035662.21  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
; OTHER INFORMATION: SWISSPROT HIT: P79955, EVALU 2.00e-22  
US-10-029-386-32842

Query Match 2.2%; Score 8; DB 12; Length 177;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
|||||  
Db 55 AGGGAGGG 62

## RESULT 6

US-10-267-430-13  
; Sequence 13, Application US/10267430  
; Publication No. US20030175311A1  
; GENERAL INFORMATION:  
; APPLICANT: Binger, Mary-Helen  
; APPLICANT: Chizzonite, Richard Anthony  
; APPLICANT: Kramer, Richard Allen  
; APPLICANT: Lomedico, Peter Thomas  
; APPLICANT: McAndrew, Stephen J.  
; APPLICANT: Altenburger, Werner  
; TITLE OF INVENTION: Recombinant Coccidioides Vaccines  
; FILE REFERENCE: Recombinant Coccidioides Vaccines  
; CURRENT APPLICATION NUMBER: US/10/267,430  
; CURRENT FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US/08/744, 793  
; PRIOR FILING DATE: 1996-11-01  
; PRIOR APPLICATION NUMBER: 07/812,349  
; PRIOR FILING DATE: 1991-12-20  
; PRIOR APPLICATION NUMBER: 07/202,721  
; PRIOR FILING DATE: 1988-06-03

; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13

; LENGTH: 306  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:

; OTHER INFORMATION: Description of Unknown Organism: AA seq. of the  
; OTHER INFORMATION: nucleotide seq. of Fig. 27.  
US-10-267-430-13

Query Match 2.2%; Score 8; DB 12; Length 306;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 NAGGGAGG 240  
|||||  
Db 206 NAGGGAGG 213

## RESULT 7

US-10-267-430-4  
; Sequence 4, Application US/10267430  
; Publication No. US20030175311A1  
; GENERAL INFORMATION:  
; APPLICANT: Binger, Mary-Helen  
; APPLICANT: Chizzonite, Richard Anthony  
; APPLICANT: Kramer, Richard Allen  
; APPLICANT: Lomedico, Peter Thomas  
; APPLICANT: McAndrew, Stephen J.  
; APPLICANT: Altenburger, Werner  
; TITLE OF INVENTION: Recombinant Coccidioides Vaccines  
; FILE REFERENCE: Recombinant Coccidioides Vaccines  
; CURRENT APPLICATION NUMBER: US/10/267,430  
; CURRENT FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US/08/744, 793  
; PRIOR FILING DATE: 1996-11-01  
; PRIOR APPLICATION NUMBER: 07/812,349  
; PRIOR FILING DATE: 1991-12-20  
; PRIOR APPLICATION NUMBER: 07/202,721  
; PRIOR FILING DATE: 1988-06-03  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: AA seq of  
; OTHER INFORMATION: nucleotide seq in Fig. 16.  
US-10-267-430-4

Query Match 2.2%; Score 8; DB 12; Length 325;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 NAGGGAGG 240  
|||||  
Db 225 NAGGGAGG 232

## RESULT 8

US-10-194-885-10  
; Sequence 10, Application US/10194885  
; Publication No. US20030135878A1  
; GENERAL INFORMATION:  
; APPLICANT: Wong, J. H.  
; APPLICANT: Cho, Myeong-Je  
; APPLICANT: Lemaux, Peggy G.  
; APPLICANT: Buchanan, Bob  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH ELEVATED  
; TITLE OF INVENTION: THIOREDUXIN LEVELS  
; FILE REFERENCE: 416272000800

; CURRENT APPLICATION NUMBER: US/10/194,885  
 ; CURRENT FILING DATE: 2002-07-12  
 ; PRIOR APPLICATION NUMBER: 60/307,006  
 ; PRIOR FILING DATE: 2001-07-19  
 ; PRIOR APPLICATION NUMBER: 09/538,864  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: 60/126,736  
 ; PRIOR FILING DATE: 1999-03-29  
 ; NUMBER OF SEQ ID NOS: 55  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 10  
 ; LENGTH: 331  
 ; TYPE: PRT  
 ; ORGANISM: Barley  
 US-10-194-885-10

Query Match 2.2%; Score 8; DB 12; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 GGAGGGPL 243  
 Db 221 GGAGGGPL 228

## RESULT 9

US-10-091-841-9  
 ; Sequence 9, Application US/10091841  
 ; Publication No. US20030150010A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cho, Myeong-Je  
 ; APPLICANT: Del Val, Greg  
 ; APPLICANT: Caillaud, Maxine  
 ; APPLICANT: Lemaux, Peggy G.  
 ; APPLICANT: Buchanan, Bob B.  
 ; TITLE OF INVENTION: Barley Gene for Thioresdoxin and  
 ; TITLE OF INVENTION: NADP-Thioresdoxin Reductase  
 ; FILE REFERENCE: 2001-0701.30  
 ; CURRENT APPLICATION NUMBER: US/10/091,841  
 ; CURRENT FILING DATE: 2002-03-05  
 ; PRIOR APPLICATION NUMBER: 09/540,014  
 ; PRIOR FILING DATE: 2000-03-31  
 ; PRIOR APPLICATION NUMBER: US 60/127,198  
 ; PRIOR FILING DATE: 1999-03-31  
 ; PRIOR APPLICATION NUMBER: US 60/169,162  
 ; PRIOR FILING DATE: 1998-12-06  
 ; PRIOR APPLICATION NUMBER: US 60/177,740  
 ; PRIOR FILING DATE: 2000-01-21  
 ; PRIOR APPLICATION NUMBER: US 60/177,739  
 ; PRIOR FILING DATE: 2000-01-21  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 9  
 ; LENGTH: 332  
 ; TYPE: PRT  
 ; ORGANISM: Hordeum vulgare  
 US-10-091-841-9

Query Match 2.2%; Score 8; DB 12; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 GGAGGGPL 243  
 Db 221 GGAGGGPL 228

## RESULT 10

US-10-295-403-148  
 ; Sequence 148, Application US/10295403  
 ; Publication No. US20030101481A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Heard, Jacqueline

; APPLICANT: Riechmann, Jose Luis  
 ; APPLICANT: Adam, Luc  
 ; APPLICANT: Broun, Pierre  
 ; APPLICANT: Pineda, Omaira  
 ; APPLICANT: Reuber, Lynne  
 ; APPLICANT: Jiang, Cai-Zhong  
 ; APPLICANT: Keddie, James  
 ; APPLICANT: Zhang, James  
 ; APPLICANT: Benito, Maria-Ines  
 ; APPLICANT: Yu, Guo-Liang  
 ; APPLICANT: Fromm, Mike  
 ; TITLE OF INVENTION: PLANT GENE SEQUENCES I  
 ; FILE REFERENCE: MBI-0003  
 ; CURRENT APPLICATION NUMBER: US/10/295,403  
 ; CURRENT FILING DATE: 2002-11-15  
 ; PRIOR APPLICATION NUMBER: US/09/394,519  
 ; PRIOR FILING DATE: 1999-09-13  
 ; PRIOR APPLICATION NUMBER: 60/101,349  
 ; PRIOR FILING DATE: 1998-09-22  
 ; PRIOR APPLICATION NUMBER: 60/103,312  
 ; PRIOR FILING DATE: 1998-10-06  
 ; PRIOR APPLICATION NUMBER: 60/108,734  
 ; PRIOR FILING DATE: 1998-11-17  
 ; PRIOR APPLICATION NUMBER: 60/113,409  
 ; PRIOR FILING DATE: 1998-12-22  
 ; NUMBER OF SEQ ID NOS: 170  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 148  
 ; LENGTH: 401  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; FEATURE:  
 ; OTHER INFORMATION: G802  
 US-10-295-403-148

Query Match 2.2%; Score 8; DB 15; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
 Db 294 AGGGAGGG 301

## RESULT 11

US-10-063-685-67  
 ; Sequence 67, Application US/10063685  
 ; Publication No. US20030180909A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Eaton, Dan L.  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Geritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Grimaldi, Christopher J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; TITLE OF INVENTION: SECATED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P3230R1C1  
 ; CURRENT APPLICATION NUMBER: US/10/063,685  
 ; CURRENT FILING DATE: 2002-05-08  
 ; PRIOR APPLICATION removed - See Palm or File Wrapper  
 ; NUMBER OF SEQ ID NOS: 170  
 ; SEQ ID NO 67  
 ; LENGTH: 744  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-063-685-67

Query Match 2.2%; Score 8; DB 12; Length 744;  
 Best Local Similarity 100.0%; Pred. No. 94;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 Db 17 AGGGAGGG 24

## RESULT 12

US-10-184-644-255  
 ; Sequence 255, Application US/10184644  
 ; Publication No. US20030044930A1

; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3430R1C227  
 ; CURRENT APPLICATION NUMBER: US/10/184,644  
 ; CURRENT FILING DATE: 2002-06-28  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 255  
 ; LENGTH: 744  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien

## US-10-184-644-255

Query Match 2.2%; Score 8; DB 15; Length 744;  
 Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 Db 17 AGGGAGGG 24

## RESULT 13

US-10-184-634-255  
 ; Sequence 255, Application US/10184634  
 ; Publication No. US20030068684A1

; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3430R1C217  
 ; CURRENT APPLICATION NUMBER: US/10/184,634  
 ; CURRENT FILING DATE: 2002-06-28  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 255  
 ; LENGTH: 744  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien

## US-10-184-634-255

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 2.2%; Score 8; DB 15; Length 744;  
 Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 Db 17 AGGGAGGG 24

## RESULT 14

US-10-140-472-153  
 ; Sequence 153, Application US/10140472  
 ; Publication No. US20030138889A1

; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C168  
 ; CURRENT APPLICATION NUMBER: US/10/140,472  
 ; CURRENT FILING DATE: 2002-05-06  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 153  
 ; LENGTH: 755  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: 67  
 ; OTHER INFORMATION: unknown base

## US-10-140-472-153

Query Match 2.2%; Score 8; DB 12; Length 755;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 Db 634 AGGGAGGG 641

## RESULT 15

US-10-141-761-153  
 ; Sequence 153, Application US/10141761  
 ; Publication No. US20030148432A1

; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood,Steven  
 ; APPLICANT: Smith,Victoria  
 ; APPLICANT: Stewart,Timothy A.  
 ; APPLICANT: Tumas,Daniel  
 ; APPLICANT: Watanabe,Colin K  
 ; APPLICANT: Wood,William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C198  
 ; CURRENT APPLICATION NUMBER: US/10/141,761  
 ; CURRENT FILING DATE: 2002-05-08  
 ; Prior Application removed - See Palm or File Wrapper  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 153  
 ; LENGTH: 755  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: 67  
 ; OTHER INFORMATION: unknown base  
 US-10-141-761-153

Query Match 2.2%; Score 8; DB 12; Length 755;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 Db 634 AGGGAGGG 641

RESULT 16  
 US-10-142-885-153  
 ; Sequence 153, Application US/10142885  
 ; Publication No. US20030157604A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C248  
 ; CURRENT APPLICATION NUMBER: US/10/142,885  
 ; CURRENT FILING DATE: 2002-05-10  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 153  
 ; LENGTH: 755  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: 67  
 ; OTHER INFORMATION: unknown base  
 US-10-142-885-153

Query Match 2.2%; Score 8; DB 12; Length 755;

Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 Db 634 AGGGAGGG 641

RESULT 17  
 US-10-158-790-153  
 ; Sequence 153, Application US/10158790  
 ; Publication No. US20030180879A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C448  
 ; CURRENT APPLICATION NUMBER: US/10/158,790  
 ; CURRENT FILING DATE: 2002-05-30  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 153  
 ; LENGTH: 755  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: 67  
 ; OTHER INFORMATION: unknown base  
 US-10-158-790-153

Query Match 2.2%; Score 8; DB 12; Length 755;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 Db 634 AGGGAGGG 641

RESULT 18  
 US-10-137-871-153  
 ; Sequence 153, Application US/10137871  
 ; Publication No. US20030207350A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria

```

; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: F3330R1C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 153
; LENGTH: 755
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 67
; OTHER INFORMATION: unknown base
; US-10-137-871-153

Query Match      2.2%; Score 8; DB 12; Length 755;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGAGGG 241
Db      634 AGGAGGG 641
|||||

RESULT 19
US-10-140-805-153
; Sequence 153, Application US/10140805
; Publication No. US20030207417A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: F3330R1C176
; CURRENT APPLICATION NUMBER: US/10/140,805
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 153
; LENGTH: 755
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 67
; OTHER INFORMATION: unknown base
; US-10-140-805-153

Query Match      2.2%; Score 8; DB 12; Length 755;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      234 AGGAGGG 241
Db      634 AGGAGGG 641
|||||

RESULT 20
US-10-140-864-153
; Sequence 153, Application US/10140864
; Publication No. US20030207419A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: F3330R1C184
; CURRENT APPLICATION NUMBER: US/10/140,864
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 153
; LENGTH: 755
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 67
; OTHER INFORMATION: unknown base
; US-10-140-864-153

Query Match      2.2%; Score 8; DB 12; Length 755;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGAGGG 241
Db      634 AGGAGGG 641
|||||

RESULT 21
US-10-140-923-153
; Sequence 153, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel

```

```
/ APPLICANT: Watanabe,Colin K
/ APPLICANT: Wood,William
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3330RIC188
/ CURRENT APPLICATION NUMBER: US/10/140,923
/ CURRENT FILING DATE: 2002-05-07
/ Prior Application removed - See Palm or File Wrapper
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 153
/ LENGTH: 755
/ TYPE: DNA
/ ORGANISM: Homo Sapien
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: 67
/ OTHER INFORMATION: unknown base
US-10-140-923-153

Query Match          2.2%; Score 8; DB 12; Length 755;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      234 AGGGAGGG 241
Db      634 AGGGAGGG 641
|||||

RESULT 22
US-10-141-756-153
/ Sequence 153, Application US/10141756
/ Publication No. US20030207359A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Beresini, Maureen
/ APPLICANT: DeForge, Laura
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerritsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3330RIC197
/ CURRENT APPLICATION NUMBER: US/10/141,759
/ CURRENT FILING DATE: 2002-05-08
/ Prior Application removed - See File Wrapper or Palm
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 153
/ LENGTH: 755
/ TYPE: DNA
/ ORGANISM: Homo Sapien
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: 67
/ OTHER INFORMATION: unknown base
US-10-141-756-153

Query Match          2.2%; Score 8; DB 12; Length 755;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      234 AGGGAGGG 241
Db      634 AGGGAGGG 641
|||||

RESULT 23
US-10-141-759-153
/ Sequence 153, Application US/10141759
/ Publication No. US20030207361A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Beresini, Maureen
/ APPLICANT: DeForge, Laura
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerritsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3330RIC197
/ CURRENT APPLICATION NUMBER: US/10/141,759
/ CURRENT FILING DATE: 2002-05-08
/ Prior Application removed - See File Wrapper or Palm
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 153
/ LENGTH: 755
/ TYPE: DNA
/ ORGANISM: Homo Sapien
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: 67
/ OTHER INFORMATION: unknown base
US-10-141-759-153

Query Match          2.2%; Score 8; DB 12; Length 755;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      234 AGGGAGGG 241
Db      634 AGGGAGGG 641
|||||

RESULT 24
US-10-123-155-153
/ Sequence 153, Application US/10123155
/ Publication No. US20030068794A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Beresini, Maureen
/ APPLICANT: DeForge, Laura
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerritsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Wood, William
```

; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P330R1C30  
; CURRENT APPLICATION NUMBER: US/10/123,155  
; CURRENT FILING DATE: 2002-04-15  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 153  
; LENGTH: 755  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 67  
; OTHER INFORMATION: unknown base  
US-10-123-155-153

Query Match 2.2%; Score 8; DB 15; Length 755;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
| | | | |  
Db 634 AGGAGGG 641

## RESULT 25

US-10-146-731-153  
; Sequence 153, Application US/10146731  
; Publication No. US20030129692A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P330R1C323  
; CURRENT APPLICATION NUMBER: US/10/146,731  
; CURRENT FILING DATE: 2002-05-15  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 153  
; LENGTH: 755  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 67  
; OTHER INFORMATION: unknown base  
US-10-146-731-153

Query Match 2.2%; Score 8; DB 16; Length 755;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
| | | | |  
Db 634 AGGAGGG 641

## RESULT 26

US-10-174-677-34  
; Sequence 34, Application US/10174677  
; Publication No. US20030190704A1  
; GENERAL INFORMATION:  
; APPLICANT: Xie, Ting  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENV  
; FILE REFERENCE: 40716 (IP-012)  
; CURRENT APPLICATION NUMBER: US/10/174,677  
; CURRENT FILING DATE: 2002-06-19  
; NUMBER OF SEQ ID NOS: 117  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 34  
; LENGTH: 828  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-174-677-34

Query Match 2.2%; Score 8; DB 12; Length 828;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
| | | | |  
Db 706 AGGAGGG 713

## RESULT 27

US-10-140-472-177  
; Sequence 177, Application US/10140472  
; Publication No. US2003013888A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P330R1C168  
; CURRENT APPLICATION NUMBER: US/10/140,472  
; CURRENT FILING DATE: 2002-05-06  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 177  
; LENGTH: 959  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-140-472-177

Query Match 2.2%; Score 8; DB 12; Length 959;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
| | | | |  
Db 676 AGGAGGG 683



```
RESULT 28
US-10-141-761-177
; Sequence 177, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; Prior Filing DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 177
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-761-177

Query Match
Best Local Similarity 2.2%; Score 8; DB 12; Length 959;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241
Db 676 AGGGAGGG 683

RESULT 29
US-10-142-885-177
; Sequence 177, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; Prior Filing DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550

Query Match
Best Local Similarity 2.2%; Score 8; DB 12; Length 959;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241
Db 676 AGGGAGGG 683

RESULT 30
US-10-158-790-177
; Sequence 177, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C448
; CURRENT APPLICATION NUMBER: US/10/158,790
; Prior Filing DATE: 2002-05-30
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 177
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-158-790-177

Query Match
Best Local Similarity 2.2%; Score 8; DB 12; Length 959;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241
Db 676 AGGGAGGG 683

RESULT 31
US-10-137-871-177
; Sequence 177, Application US/10137871
; Publication No. US20030207350A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
```

```
; SEQ ID NO 177
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-885-177

Query Match
Best Local Similarity 2.2%; Score 8; DB 12; Length 959;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241
Db 676 AGGGAGGG 683

RESULT 32
US-10-137-871-177
; Sequence 177, Application US/10137871
; Publication No. US20030207350A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
```

```

; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; Prior Application removed - See Palm or File Wrapper
; SEQ ID NO 177
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-137-871-177

Query Match      2.2%; Score 8; DB 12; Length 959;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGGAGGG 241
Db      676 AGGGAGGG 683
|||||

RESULT 32
US-10-140-805-177
; Sequence 177, Application US/10140805
; Publication No. US20030207417A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C176
; CURRENT APPLICATION NUMBER: US/10/140,805
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; SEQ ID NO 177
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-805-177

Query Match      2.2%; Score 8; DB 12; Length 959;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGGAGGG 241
Db      676 AGGGAGGG 683
|||||

US-10-140-805-177
; Sequence 177, Application US/10140805
; Publication No. US20030207417A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C176
; CURRENT APPLICATION NUMBER: US/10/140,805
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; SEQ ID NO 177
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-805-177

Query Match      2.2%; Score 8; DB 12; Length 959;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGGAGGG 241
Db      676 AGGGAGGG 683
|||||

```

```

RESULT 33
US-10-140-864-177
; Sequence 177, Application US/10140864
; Publication No. US20030207419A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C184
; CURRENT APPLICATION NUMBER: US/10/140,864
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; SEQ ID NO 177
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-864-177

Query Match      2.2%; Score 8; DB 12; Length 959;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGGAGGG 241
Db      676 AGGGAGGG 683
|||||

RESULT 34
US-10-140-923-177
; Sequence 177, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C188
; CURRENT APPLICATION NUMBER: US/10/140,923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper

```

; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 177  
 ; LENGTH: 959  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-140-923-177

Query Match 2.2%; Score 8; DB 12; Length 959;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
 |||||  
 Db 676 AGGGAGGG 683

## RESULT 35

US-10-141-756-177  
 ; Sequence 177, Application US/10141756  
 ; Publication No. US20030207359A1  
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C200  
 ; CURRENT APPLICATION NUMBER: US/10/141,756  
 ; CURRENT FILING DATE: 2002-05-08  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 177  
 ; LENGTH: 959  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-141-756-177

Query Match 2.2%; Score 8; DB 12; Length 959;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
 |||||  
 Db 676 AGGGAGGG 683

## RESULT 36

US-10-141-759-177  
 ; Sequence 177, Application US/10141759  
 ; Publication No. US20030207361A1  
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C197  
 ; CURRENT APPLICATION NUMBER: US/10/141,759  
 ; CURRENT FILING DATE: 2002-05-08  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 177  
 ; LENGTH: 959  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-141-759-177

Query Match 2.2%; Score 8; DB 12; Length 959;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
 |||||  
 Db 676 AGGGAGGG 683

## RESULT 37

US-10-123-155-177  
 ; Sequence 177, Application US/10123155  
 ; Publication No. US20030068794A1  
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C30  
 ; CURRENT APPLICATION NUMBER: US/10/123,155  
 ; CURRENT FILING DATE: 2002-04-15  
 ; Prior Application removed - See Palm or File Wrapper  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 177  
 ; LENGTH: 959  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-123-155-177

Query Match 2.2%; Score 8; DB 15; Length 959;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGGG 241  
 |||||  
 Db 676 AGGGAGGG 683

## RESULT 38

US-10-146-731-177  
 ; Sequence 177, Application US/10146731  
 ; Publication No. US20030129692A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C323  
 ; CURRENT APPLICATION NUMBER: US/10/146,731  
 ; CURRENT FILING DATE: 2002-05-15  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 177  
 ; LENGTH: 959  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-146-731-177

Query Match 2.2%; Score 8; DB 16; Length 959;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 Db 676 AGGGAGGG 683

## RESULT 39

US-10-184-644-23  
 ; Sequence 23, Application US/10184644  
 ; Publication No. US20030044930A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3430R1C227  
 ; CURRENT APPLICATION NUMBER: US/10/184,644  
 ; CURRENT FILING DATE: 2002-06-28  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 23  
 ; LENGTH: 997  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien

## US-10-184-644-23

Query Match 2.2%; Score 8; DB 15; Length 997;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 Db 51 AGGGAGGG 58

## RESULT 40

US-10-184-634-23  
 ; Sequence 23, Application US/10184634  
 ; Publication No. US2003006864A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3430R1C217  
 ; CURRENT APPLICATION NUMBER: US/10/184,634  
 ; CURRENT FILING DATE: 2002-06-28  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 23  
 ; LENGTH: 997  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-184-634-23

Query Match 2.2%; Score 8; DB 15; Length 997;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 Db 51 AGGGAGGG 58

## RESULT 41

US-09-738-626-5676  
 ; Sequence 5676, Application US/09738626  
 ; Publication No. US20020197605A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAGAWA, SATOSHI  
 ; APPLICANT: MIZOUCHI, HIROSHI  
 ; APPLICANT: ANDO, SEIKO  
 ; APPLICANT: HAYASHI, MIKIRO  
 ; APPLICANT: OCHIAI, KEIKO  
 ; APPLICANT: YOKOI, HARUHIKO  
 ; APPLICANT: TATEISHI, NAOKO  
 ; APPLICANT: SENO, AKIHIRO  
 ; APPLICANT: IKEDA, MASATO  
 ; APPLICANT: OZAKI, AKIO  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-125  
 ; CURRENT APPLICATION NUMBER: US/09/738,626  
 ; CURRENT FILING DATE: 2000-12-18  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 997  
 ; SEQ ID NO 23  
 ; LENGTH: 997  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien

; PRIOR FILING DATE: 2000-08-03  
 ; NUMBER OF SEQ ID NOS: 7059  
 ; SOFTWARE: PatentIn ver. 3.0  
 ; SEQ ID NO 5676  
 ; LENGTH: 1004  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium glutamicum  
 US-09-738-626-5676

Query Match 2.2%; Score 8; DB 10; Length 1004;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
 |||||  
 Db 338 AGGAGGG 345

## RESULT 42

US-10-184-644-99  
 ; Sequence 99, Application US/10184644  
 ; Publication No. US20030044930A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P3430R1C227  
 ; CURRENT APPLICATION NUMBER: US/10/184,644  
 ; CURRENT FILING DATE: 2002-06-28  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 99  
 ; LENGTH: 1016  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-184-644-99

Query Match 2.2%; Score 8; DB 15; Length 1016;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
 |||||  
 Db 667 AGGAGGG 674

## RESULT 43

US-10-184-634-99  
 ; Sequence 99, Application US/10184634  
 ; Publication No. US20030068684A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P3430R1C217  
 ; CURRENT APPLICATION NUMBER: US/10/184,634  
 ; CURRENT FILING DATE: 2002-06-28  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 99  
 ; LENGTH: 1016  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-184-634-99

Query Match 2.2%; Score 8; DB 15; Length 1016;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
 |||||  
 Db 667 AGGAGGG 674

## RESULT 44

US-10-063-685-87  
 ; Sequence 87, Application US/10063685  
 ; Publication No. US20030180909A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Eaton, Dan L.  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Grimaldi, Christopher J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P3230R1C1  
 ; CURRENT APPLICATION NUMBER: US/10/063,685  
 ; CURRENT FILING DATE: 2002-05-08  
 ; Prior Application removed - See Palm or File Wrapper  
 ; NUMBER OF SEQ ID NOS: 170  
 ; SEQ ID NO 87  
 ; LENGTH: 1176  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-063-685-87

Query Match 2.2%; Score 8; DB 12; Length 1176;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
 |||||  
 Db 38 AGGAGGG 45

## RESULT 45

US-10-184-644-293  
 ; Sequence 293, Application US/10184644  
 ; Publication No. US20030044930A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin

```
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 293
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-293

Query Match      2.2%; Score 8; DB 15; Length 1176;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      234 AGGAGGG 241
Db      38 AGGAGGG 45

RESULT 46
US-10-184-634-293
; Sequence 293, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 293
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-293

Query Match      2.2%; Score 8; DB 15; Length 1176;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      234 AGGAGGG 241
Db      38 AGGAGGG 45

RESULT 47
US-09-826-508-3
; Sequence 3, Application US/09826508
; Patent No. US20010025099A1
; GENERAL INFORMATION:
; APPLICANT: Nabil Elshourbagy
; APPLICANT: Lisa Vawter
; TITLE OF INVENTION: G Protein-Coupled Receptor Polypeptides
; FILE REFERENCE: GP-70744USB
; CURRENT APPLICATION NUMBER: US/09/826,508
; CURRENT FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 40
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1200
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-826-508-3

Query Match      2.2%; Score 8; DB 9; Length 1200;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      234 AGGAGGG 241
Db      1087 AGGAGGG 1094

RESULT 48
US-10-184-644-223
; Sequence 223, Application US/10184644
; Publication No. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 223
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-223

Query Match      2.2%; Score 8; DB 15; Length 1245;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      234 AGGAGGG 241
Db      38 AGGAGGG 45

RESULT 49
US-10-184-634-223
; Sequence 223, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
```

; CURRENT APPLICATION NUMBER: US/10/184,634  
; CURRENT FILING DATE: 2002-06-28  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 223  
; LENGTH: 1245  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-184-634-223

Query Match 2.2%; Score 8; DB 15; Length 1245;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
Db 38 AGGAGGG 45  
|||||

RESULT 50  
US-10-356-153-86  
; Sequence 86, Application US/10356153  
; Publication No. US20030166176A1  
; GENERAL INFORMATION:  
; APPLICANT: Croteau, Rodney et al.  
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES  
; FILE REFERENCE: 62773  
; CURRENT APPLICATION NUMBER: US/10/356,153  
; CURRENT FILING DATE: 2003-01-29  
; PRIOR APPLICATION NUMBER: US/10/142,231  
; PRIOR FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 60/165,250  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 86  
; LENGTH: 1461  
; TYPE: PRT  
; ORGANISM: Taxus cuspidata  
US-10-356-153-86

Query Match 2.2%; Score 8; DB 12; Length 1461;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
Db 817 AGGAGGG 824  
|||||

RESULT 51  
US-10-142-231-86  
; Sequence 86, Application US/10142231  
; Publication No. US2003007796A1  
; GENERAL INFORMATION:  
; APPLICANT: Croteau, Rodney et al.  
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES  
; FILE REFERENCE: 62773  
; CURRENT APPLICATION NUMBER: US/10/142,231  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 60/165,250  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 86  
; LENGTH: 1461  
; TYPE: PRT  
; ORGANISM: Taxus cuspidata  
US-10-142-231-86

Query Match 2.2%; Score 8; DB 15; Length 1461;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
Db 817 AGGAGGG 824  
|||||

RESULT 52  
US-10-140-472-363  
; Sequence 363, Application US/10140472  
; Publication No. US2003013888A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C168  
; CURRENT APPLICATION NUMBER: US/10/140,472  
; CURRENT FILING DATE: 2002-05-06  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 363  
; LENGTH: 1503  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-140-472-363

Query Match 2.2%; Score 8; DB 12; Length 1503;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
Db 1203 AGGAGGG 1210  
|||||

RESULT 53  
US-10-141-761-363  
; Sequence 363, Application US/10141761  
; Publication No. US20030148432A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: P3330R1C198  
 ; CURRENT APPLICATION NUMBER: US/10/141,761  
 ; CURRENT FILING DATE: 2002-05-08  
 ; Prior Application removed - See Palm or File Wrapper  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 363  
 ; LENGTH: 1503  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-141-761-363

Query Match 2.2%; Score 8; DB 12; Length 1503;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 |||||  
 Db 1203 AGGGAGGG 1210

RESULT 54  
 US-10-142-885-363  
 ; Sequence 363, Application US/10142885  
 ; Publication No. US20030157604A1  
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C248  
 ; CURRENT APPLICATION NUMBER: US/10/142,885

; CURRENT FILING DATE: 2002-05-10  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 363

; LENGTH: 1503  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-142-885-363

Query Match 2.2%; Score 8; DB 12; Length 1503;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 |||||  
 Db 1203 AGGGAGGG 1210

RESULT 55  
 US-10-158-790-363  
 ; Sequence 363, Application US/10158790  
 ; Publication No. US20030180879A1  
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C448  
 ; CURRENT APPLICATION NUMBER: US/10/158,790  
 ; CURRENT FILING DATE: 2002-05-30  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 363  
 ; LENGTH: 1503  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-158-790-363

Query Match 2.2%; Score 8; DB 12; Length 1503;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 |||||  
 Db 1203 AGGGAGGG 1210

RESULT 56  
 US-10-137-871-363  
 ; Sequence 363, Application US/10137871  
 ; Publication No. US20030207350A1  
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C153  
 ; CURRENT APPLICATION NUMBER: US/10/137,871  
 ; CURRENT FILING DATE: 2002-05-03

; Prior Application removed - See Palm or File Wrapper  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 363  
 ; LENGTH: 1503  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-137-871-363

Query Match 2.2%; Score 8; DB 12; Length 1503;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;



Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||||  
Db 1203 AGGAGGG 1210

## RESULT 57

US-10-140-805-363  
; Sequence 363, Application US/10140805  
; Publication No. US20030207417A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330R1C176  
; CURRENT APPLICATION NUMBER: US/10/140,805  
; CURRENT FILING DATE: 2002-05-07  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 363  
; LENGTH: 1503  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-140-805-363

Query Match 2.2%; Score 8; DB 12; Length 1503;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||||  
Db 1203 AGGAGGG 1210

## RESULT 58

US-10-140-864-363  
; Sequence 363, Application US/10140864  
; Publication No. US20030207419A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330R1C184  
; CURRENT APPLICATION NUMBER: US/10/140,864  
; CURRENT FILING DATE: 2002-05-07  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 363  
; LENGTH: 1503  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-140-864-363

Query Match 2.2%; Score 8; DB 12; Length 1503;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||||  
Db 1203 AGGAGGG 1210

## RESULT 59

US-10-140-923-363  
; Sequence 363, Application US/10140923  
; Publication No. US20030207355A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330R1C188  
; CURRENT APPLICATION NUMBER: US/10/140,923  
; CURRENT FILING DATE: 2002-05-07  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 363  
; LENGTH: 1503  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-140-923-363

Query Match 2.2%; Score 8; DB 12; Length 1503;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
|||||||  
Db 1203 AGGAGGG 1210

## RESULT 60

US-10-141-756-363  
; Sequence 363, Application US/10141756  
; Publication No. US20030207359A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge,Laura  
 ; APPLICANT: Desnoyers,Luc  
 ; APPLICANT: Filvaroff,Ellen  
 ; APPLICANT: Gao,Wei-Qiang  
 ; APPLICANT: Gerritsen,Mary E.  
 ; APPLICANT: Goddard,Audrey  
 ; APPLICANT: Godowski,Paul J.  
 ; APPLICANT: Gurney,Austin L.  
 ; APPLICANT: Sherwood,Steven  
 ; APPLICANT: Smith,Victoria  
 ; APPLICANT: Stewart,Timothy A.  
 ; APPLICANT: Tumas,Daniel  
 ; APPLICANT: Watanabe,Colin K  
 ; APPLICANT: Wood,William  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C200  
 ; CURRENT APPLICATION NUMBER: US/10/141,756  
 ; CURRENT FILING DATE: 2002-05-08  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 363  
 ; LENGTH: 1503  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-10-141-756-363

Query Match 2.2%; Score 8; DB 12; Length 1503;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 |||||  
 DB 1203 AGGGAGGG 1210

## RESULT 61

US-10-141-759-363  
 ; Sequence 363, Application US/10141759  
 ; Publication No. US20030207361A1  
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C197  
 ; CURRENT APPLICATION NUMBER: US/10/141,759  
 ; CURRENT FILING DATE: 2002-05-08  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 550  
 ; SEQ ID NO 363

LENGTH: 1503

TYPE: DNA

ORGANISM: Homo Sapien

US-10-141-759-363

Query Match 2.2%; Score 8; DB 12; Length 1503;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 |||||  
 DB 1203 AGGGAGGG 1210

## RESULT 62

US-10-123-155-363  
 ; Sequence 363, Application US/10123155  
 ; Publication No. US20030068794A1  
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3330R1C30  
 ; CURRENT APPLICATION NUMBER: US/10/123,155

FILE REFERENCE: P3330R1C30

CURRENT APPLICATION NUMBER: US/10/123,155

CURRENT FILING DATE: 2002-04-15

Prior Application removed - See Palm or File Wrapper

NUMBER OF SEQ ID NOS: 550

SEQ ID NO 363

LENGTH: 1503

TYPE: DNA

ORGANISM: Homo Sapien

US-10-123-155-363

Query Match 2.2%; Score 8; DB 15; Length 1503;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241  
 |||||  
 DB 1203 AGGGAGGG 1210

## RESULT 63

US-10-146-731-363  
 ; Sequence 363, Application US/10146731  
 ; Publication No. US20030129692A1  
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerritsen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William

```

; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C323
; CURRENT APPLICATION NUMBER: US/10/146,731
; CURRENT FILING DATE: 2002-05-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 363
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-146-731-363

Query Match      2.2%; Score 8; DB 16; Length 1503;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGGAGGG 241
Db      1203 AGGGAGGG 1210

RESULT 64
US-10-184-644-421
; Sequence 421, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 421
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-421

Query Match      2.2%; Score 8; DB 15; Length 1524;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGGAGGG 241
Db      923 AGGGAGGG 930

RESULT 65
US-10-184-634-421
; Sequence 421, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James

```

```

; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 421
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-421

Query Match      2.2%; Score 8; DB 15; Length 1524;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGGAGGG 241
Db      923 AGGGAGGG 930

RESULT 66
US-10-184-644-245
; Sequence 245, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 245
; LENGTH: 1564
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-245

Query Match      2.2%; Score 8; DB 15; Length 1564;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGGAGGG 241
Db      903 AGGGAGGG 910

RESULT 67
US-10-184-634-245
; Sequence 245, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey

```

```

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; PRIOR FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 245
; LENGTH: 1564
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-245

Query Match      2.2%; Score 8; DB 15; Length 1564;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      234 AGGGAGGG 241
Db      903 AGGGAGGG 910

RESULT 68
US-10-140-472-373
; Sequence 373, Application US/10140472
; Publication No. US2003013888A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; PRIOR FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 373
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-761-373

Query Match      2.2%; Score 8; DB 12; Length 1572;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      234 AGGGAGGG 241
Db      709 AGGGAGGG 716

RESULT 70
US-10-142-885-373
; Sequence 373, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550

```

```

RESULT 69
US-10-141-761-373
; Sequence 373, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; PRIOR FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 373
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-761-373

Query Match      2.2%; Score 8; DB 12; Length 1572;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      234 AGGGAGGG 241
Db      709 AGGGAGGG 716

RESULT 70
US-10-142-885-373
; Sequence 373, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550

```

; SEQ ID NO 373  
; LENGTH: 1572  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-142-885-373

Query Match 2.2%; Score 8; DB 12; Length 1572;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
Db 709 AGGAGGG 716

## RESULT 71

US-10-158-790-373  
; Sequence 373, Application US/10158790  
; Publication No. US20030180879A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C448  
; CURRENT APPLICATION NUMBER: US/10/158,790  
; CURRENT FILING DATE: 2002-05-30  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 373  
; LENGTH: 1572  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-158-790-373

Query Match 2.2%; Score 8; DB 12; Length 1572;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
Db 709 AGGAGGG 716

## RESULT 72

US-10-137-871-373  
; Sequence 373, Application US/10137871  
; Publication No. US20030207350A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C153  
; CURRENT APPLICATION NUMBER: US/10/137,871  
; CURRENT FILING DATE: 2002-05-03  
; Prior Application removed - See File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 373  
; LENGTH: 1572  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-137-871-373

Query Match 2.2%; Score 8; DB 12; Length 1572;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
Db 709 AGGAGGG 716

## RESULT 73

US-10-140-805-373  
; Sequence 373, Application US/10140805  
; Publication No. US20030207417A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C176  
; CURRENT APPLICATION NUMBER: US/10/140,805  
; CURRENT FILING DATE: 2002-05-07  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 373  
; LENGTH: 1572  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-140-805-373

Query Match 2.2%; Score 8; DB 12; Length 1572;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGGG 241  
Db 709 AGGAGGG 716

## RESULT 74

US-10-140-864-373  
; Sequence 373, Application US/10140864  
; Publication No. US20030207419A1  
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Smith, Victoria  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C184

; CURRENT APPLICATION NUMBER: US/10/140,864

; CURRENT FILING DATE: 2002-05-07

; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 373

; LENGTH: 1572

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-140-864-373

Query Match

Best Local Similarity 2.2%; Score 8; DB 12; Length 1572;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241

Db 709 AGGGAGGG 716

## RESULT 75

US-10-140-923-373

; Sequence 373, Application US/10140923

; Publication No. US20030207355A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Smith, Victoria  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C188

; CURRENT APPLICATION NUMBER: US/10/140,923

; CURRENT FILING DATE: 2002-05-07

; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 373

; LENGTH: 1572

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-140-923-373

Query Match

Best Local Similarity 2.2%; Score 8; DB 12; Length 1572;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGGAGGG 241

Db 709 AGGGAGGG 716

Search completed: December 17, 2003, 06:43:01

Job time : 32.9444 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Kun on: December 17, 2003, 06:32:02 ; Search time 12.8538 Seconds  
(without alignments)  
1211.344 Million cell updates/sec

Title: US-09-805-550-4

Perfect score: 368

Sequence: 1 MKLTVKTLKGTHFEIRVQPN.....CORNEELAANYLLEHAGED 368

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PTUS-COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfilesi.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	368	100.0	368	3	US-09-413-574-4
2	13	3.5	405	3	US-09-413-574-2
3	8	2.2	332	4	US-09-540-014-9
4	8	2.2	371	4	US-09-252-991A-25766
5	8	2.2	379	4	US-09-252-991A-28428
6	8	2.2	738	3	US-08-864-038A-3
7	8	2.2	1417	4	US-08-900-230-3
8	7	1.9	20	3	US-08-961-308B-5
9	7	1.9	57	3	US-08-506-553C-1
10	7	1.9	61	4	US-09-149-476-615
11	7	1.9	81	4	US-09-249-542-9
12	7	1.9	82	4	US-09-249-542-8
13	7	1.9	96	3	US-09-036-113-2
14	7	1.9	96	4	US-09-556-166-2
15	7	1.9	96	4	US-09-252-991A-29877
16	7	1.9	111	4	US-09-249-542-16
17	7	1.9	111	4	US-09-249-542-17
18	7	1.9	111	4	US-09-249-542-18
19	7	1.9	112	4	US-09-249-542-7
20	7	1.9	112	4	US-09-249-542-10
21	7	1.9	112	4	US-09-249-542-11
22	7	1.9	112	4	US-09-249-542-14
23	7	1.9	112	4	US-09-249-542-15
24	7	1.9	112	4	US-09-249-542-19
25	7	1.9	112	4	US-09-249-542-24
26	7	1.9	128	4	US-09-996-243-420
27	7	1.9	160	4	US-09-328-352-6723

28	7	1.9	161	2	US-08-403-852D-30	Sequence 30, Appl
29	7	1.9	161	3	US-08-510-646B-31	Sequence 31, Appl
30	7	1.9	161	3	US-09-231-818-30	Sequence 30, Appl
31	7	1.9	185	4	US-09-562-737-112	Sequence 112, Appl
32	7	1.9	208	2	US-08-606-143-40	Sequence 40, Appl
33	7	1.9	215	4	US-08-252-991A-25666	Sequence 25666, A
34	7	1.9	216	3	US-08-506-553C-23	Sequence 23, Appl
35	7	1.9	226	4	US-09-107-532A-5362	Sequence 5362, Ap
36	7	1.9	235	2	US-08-529-190B-1	Sequence 1, Appl
37	7	1.9	238	4	US-09-219-983A-12	Sequence 12, Appl
38	7	1.9	240	1	US-08-414-926A-10	Sequence 10, Appl
39	7	1.9	240	2	US-08-926-922-10	Sequence 10, Appl
40	7	1.9	240	3	US-09-253-682-10	Sequence 10, Appl
41	7	1.9	240	3	US-09-527-657-10	Sequence 10, Appl
42	7	1.9	242	4	US-09-252-991A-22027	Sequence 22027, A
43	7	1.9	254	4	US-09-266-965-123	Sequence 123, Appl
44	7	1.9	267	4	US-09-219-983A-11	Sequence 11, Appl
45	7	1.9	269	4	US-08-252-991A-16773	Sequence 16773, A
46	7	1.9	276	3	US-08-506-553C-26	Sequence 26, Appl
47	7	1.9	281	4	US-09-252-991A-19858	Sequence 19858, A
48	7	1.9	307	4	US-09-252-991A-26854	Sequence 26854, A
49	7	1.9	313	4	US-09-252-991A-22411	Sequence 22411, A
50	7	1.9	322	4	US-09-252-991A-22018	Sequence 22018, A
51	7	1.9	323	4	US-09-252-991A-33113	Sequence 33113, A
52	7	1.9	340	4	US-09-328-352-5177	Sequence 5177, Ap
53	7	1.9	352	4	US-09-252-991A-17906	Sequence 17906, A
54	7	1.9	366	4	US-09-328-352-5630	Sequence 23504, A
55	7	1.9	372	4	US-09-328-352-5630	Sequence 17906, A
56	7	1.9	441	4	US-09-252-991A-24396	Sequence 24396, A
57	7	1.9	486	2	US-08-821-355A-8	Sequence 8, Appl
58	7	1.9	486	2	US-09-003-687A-8	Sequence 8, Appl
59	7	1.9	486	3	US-09-136-605-8	Sequence 8, Appl
60	7	1.9	511	2	US-08-821-355A-9	Sequence 9, Appl
61	7	1.9	511	2	US-09-003-687A-9	Sequence 9, Appl
62	7	1.9	511	3	US-08-136-605-9	Sequence 9, Appl
63	7	1.9	521	4	US-09-252-991A-18266	Sequence 18266, A
64	7	1.9	539	3	US-08-687-590-27	Sequence 27, Appl
65	7	1.9	539	3	US-09-702-705-326	Sequence 326, App
66	7	1.9	539	4	US-09-736-457-326	Sequence 326, App
67	7	1.9	546	2	US-09-067-351-1	Sequence 1, Appl
68	7	1.9	546	3	US-09-360-490-1	Sequence 1, Appl
69	7	1.9	590	3	US-09-413-814-89	Sequence 89, Appl
70	7	1.9	591	3	US-09-413-814-76	Sequence 76, Appl
71	7	1.9	609	4	US-09-252-991A-20134	Sequence 20134, A
72	7	1.9	612	4	US-08-894-454-163	Sequence 163, App
73	7	1.9	615	4	US-09-328-352-4474	Sequence 4474, Ap
74	7	1.9	627	4	US-09-345-473E-46	Sequence 46, Appl
75	7	1.9	627	4	US-09-345-473E-47	Sequence 47, Appl
76	7	1.9	628	4	US-09-345-473E-48	Sequence 48, Appl
77	7	1.9	641	4	US-09-249-585A-3	Sequence 3, Appl
78	7	1.9	641	4	US-09-410-399-4	Sequence 4, Appl
79	7	1.9	675	1	US-08-386-495-10	Sequence 10, Appl
80	7	1.9	675	5	PCT-US96-02331-10	Sequence 10, Appl
81	7	1.9	678	4	US-09-107-532A-5685	Sequence 5685, Ap
82	7	1.9	728	4	US-09-252-991A-31891	Sequence 31891, A
83	7	1.9	753	4	US-09-252-991A-17612	Sequence 17612, A
84	7	1.9	800	1	US-08-785-052-4	Sequence 4, Appl
85	7	1.9	800	2	US-08-913-581-4	Sequence 4, Appl
86	7	1.9	801	1	US-07-906-349A-6	Sequence 6, Appl
87	7	1.9	887	4	US-09-077-940A-2	Sequence 2, Appl
88	7	1.9	924	1	US-08-481-130-28	Sequence 28, Appl
89	7	1.9	924	1	US-08-656-984A-28	Sequence 28, Appl
90	7	1.9	924	1	US-08-485-604-28	Sequence 28, Appl
91	7	1.9	924	2	US-08-487-595-28	Sequence 28, Appl
92	7	1.9	971	2	US-08-724-354D-22	Sequence 22, Appl
93	7	1.9	971	3	US-09-270-984A-22	Sequence 22, Appl
94	7	1.9	971	3	US-09-177-431-8	Sequence 8, Appl
95	7	1.9	988	4	US-09-252-991A-29699	Sequence 29699, A
96	7	1.9	1000	4	US-09-252-991A-31361	Sequence 31361, A
97	7	1.9	1075	4	US-09-252-991A-18387	Sequence 18387, A
98	7	1.9	1093	5	PCT-US94-0449D-55	Sequence 55, Appl
99	7	1.9	1093	5	PCT-US94-0449D-55	Sequence 55, Appl
100	7	1.9	1323	1	US-08-026-138E-4	Sequence 4, Appl

## ALIGNMENTS

```
RESULT 1
US-09-413-574-4
; Sequence 4, Application US/09413574
; Patent No. 6235972
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Tagliani, Laura
; TITLE OF INVENTION: Maize Rad23 Genes and Uses Thereof
; FILE REFERENCE: 0964
; CURRENT APPLICATION NUMBER: US/09/413,574
; CURRENT FILING DATE: 1999-10-06
; EARLIER APPLICATION NUMBER: 60/109,728
; EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-09-413-574-4

Query Match
Best Local Similarity 100.0%; Score 368; DB 3; Length 368;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLTVKLTGTHFEIRVQPNNDTIMAVKKNIEEIQGKDSYPMGQQLIFNGKVLKDESTLE 60
DB 1 MKLTVKLTGTHFEIRVQPNNDTIMAVKKNIEEIQGKDSYPMGQQLIFNGKVLKDESTLE 60

QY 61 ENKVNEDGFLVVMLSKGKTSSTGTSQSHSNTPATRQAPPLPAPQAPQPPVAPITTSQ 120
DB 61 ENKVNEDGFLVVMLSKGKTSSTGTSQSHSNTPATRQAPPLPAPQAPQPPVAPITTSQ 120

QY 121 PEGLPAAQPNTHNAASNLSSGRNVDITINQLMEMGGGSKDKVORALRAAYNNPERAV 180
DB 121 PEGLPAAQPNTHNAASNLSSGRNVDITINQLMEMGGGSKDKVORALRAAYNNPERAV 180

QY 181 EYLYSGIPVTAETAVPIGGGANTTDRAPTGEAGLSGIPNTAPLDLFPQASNAGGAGG 240
DB 181 EYLYSGIPVTAETAVPIGGGANTTDRAPTGEAGLSGIPNTAPLDLFPQASNAGGAGG 240

QY 241 GPLDFLRNNPQFQAVREMVHTNPQILQPMVLVELSKQNPQILRLIEENHDEFLQLNEPPE 300
DB 241 GPLDFLRNNPQFQAVREMVHTNPQILQPMVLVELSKQNPQILRLIEENHDEFLQLNEPPE 300

QY 301 GGGGDFLDQPEDEMPHAI SVTPEQEAIGRLSMGDFDRARVIEAFACDRNEELAAANYL 360
DB 301 GGGGDFLDQPEDEMPHAI SVTPEQEAIGRLSMGDFDRARVIEAFACDRNEELAAANYL 360

QY 361 LEHAGEED 368
DB 361 LEHAGEED 368

RESULT 2
US-09-413-574-2
; Sequence 2, Application US/09413574
; Patent No. 6235972
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Tagliani, Laura
; TITLE OF INVENTION: Maize Rad23 Genes and Uses Thereof
; FILE REFERENCE: 0964
; CURRENT APPLICATION NUMBER: US/09/413,574
; CURRENT FILING DATE: 1999-10-06
; EARLIER APPLICATION NUMBER: 60/109,728
; EARLIER FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Zea mays
US-09-413-574-2

Query Match
Best Local Similarity 100.0%; Score 13; DB 3; Length 405;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 RALRAAYNNPERA 179
DB 162 RALRAAYNNPERA 194

RESULT 3
US-09-540-014-9
; Sequence 9, Application US/09540014
; Patent No. 6380372
; GENERAL INFORMATION:
; APPLICANT: Cho, Myeong-Je
; APPLICANT: Del Val, Greg
; APPLICANT: Caillaud, Maxime
; APPLICANT: Lemauz, Peggy G.
; APPLICANT: Buchanan, Bob B.
; TITLE OF INVENTION: Barley Gene for Thioredoxin and
; TITLE OF INVENTION: NADP-Thioredoxin Reductase
; FILE REFERENCE: 2001-0701.30
; CURRENT APPLICATION NUMBER: US/09/540,014
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: US 60/127,198
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: US 60/169,162
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: US 60/177,740
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 60/177,739
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-09-540-014-9

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 332;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGPL 243
DB 221 GGAGGGPL 228

RESULT 4
US-09-252-991A-25766
; Sequence 25766, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25766
```



```
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25766

Query Match          2.2%; Score 8; DB 4; Length 371;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 PEGLPQA 128
Db 151 PEGLPQA 158

RESULT 5
US-09-252-991A-28428
; Sequence 28428, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28428
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28428

Query Match          2.2%; Score 8; DB 4; Length 379;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 NAGGGAGG 240
Db 217 NAGGGAGG 224

RESULT 6
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE cDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isehinden
; CITY: Teu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: B (by experiment)
US-08-864-038A-3

Query Match          2.2%; Score 8; DB 3; Length 738;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGG 241
Db 119 AGGGAGG 126

RESULT 7
US-08-900-230-3
; Sequence 3, Application US/08900230
; Patent No. 6329197
; GENERAL INFORMATION:
; APPLICANT: Bard, Jonathan A.
; TITLE OF INVENTION: DNA ENCODING GALANN GALR3 RECEPTORS AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of The Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,230
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52241-C/JPW/ADM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1417 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHETICAL: NO
```

ANTI-SENSE: NO  
US-08-900-230-3

Query Match 2.2%; Score 8; DB 4; Length 1417;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 241  
Db 1237 AGGAGGG 1244

# RESULT 8

US-08-961-308B-5  
Sequence 5, Application US/08961308B

Patent No. 6248539

GENERAL INFORMATION:

APPLICANT: Ghadiri, Reza M.

APPLICANT: Sailor, Michael J.

APPLICANT: Motesharei, Kianoush

APPLICANT: Lin, Shang-Yi

APPLICANT: Dancil, Keiki-Pua S.

TITLE OF INVENTION: A Porous Silicon-Based Optical Interferometric Sensor

FILE REFERENCE: 4583-71578

CURRENT APPLICATION NUMBER: US/08/961,308B

CURRENT FILING DATE: 1997-10-30

PRIOR APPLICATION NUMBER: 08/924,601

PRIOR FILING DATE: 1997-09-05

NUMBER OF SEQ ID NOS: 5

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5

LENGTH: 20

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

OTHER INFORMATION: polypeptide related to Epstein-Barr Virus Nuclear

OTHER INFORMATION: Antigen

US-08-961-308B-5

# Query Match

Best Local Similarity 1.9%; Score 7; DB 3; Length 20;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGGG 240  
Db 3 AGGAGGG 9

# RESULT 9

US-08-506-553C-1

Sequence 1, Application US/08506553C

Patent No. 6120989

GENERAL INFORMATION:

APPLICANT: Vornhagen, Rolf; Hinderer, Walter; Sonneborn, Han-H.

APPLICANT: Plachter, Bodo; and Jahn, Gerhard

TITLE OF INVENTION: ISOLATED HUMAN CYTOMEGALOVIRUS

TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felfe & Lynch

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/506,553C

SEQUENCE 1, Application US/09149476

Patent No. 6420526

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: 186 Human Secreted proteins

FILE REFERENCE: P2002P1

CURRENT APPLICATION NUMBER: US/09/149,476

CURRENT FILING DATE: 1998-09-08

EARLIER APPLICATION NUMBER: PCT/US98/04493

EARLIER FILING DATE: 1998-03-06

EARLIER APPLICATION NUMBER: 60/040,162

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,333

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/038,621

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,626

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,334

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,336

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,163

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/047,600

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,615

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,597

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,502

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,633

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,583

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,617

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,618

EARLIER FILING DATE: 1997-05-23

FILING DATE: 07/25/95  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary Anne Schofield  
REGISTRATION NUMBER: 36,669  
REFERENCE/DOCKET NUMBER: LEADER 202-PPF/MAS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 57 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Protein

FRAGMENT TYPE: internal

US-08-506-553C-1

Query Match

Best Local Similarity 1.9%; Score 7; DB 3; Length 57;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
Db 4 GGGAGGG 10

# RESULT 10

US-09-149-476-615

Sequence 615, Application US/09149476

Patent No. 6420526

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: 186 Human Secreted proteins

FILE REFERENCE: P2002P1

CURRENT APPLICATION NUMBER: US/09/149,476

CURRENT FILING DATE: 1998-09-08

EARLIER APPLICATION NUMBER: PCT/US98/04493

EARLIER FILING DATE: 1998-03-06

EARLIER APPLICATION NUMBER: 60/040,162

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,333

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/038,621

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,626

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,334

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,336

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,163

EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/047,600

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,615

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,597

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,502

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,633

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,583

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,617

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,618

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,618

EARLIER FILING DATE: 1997-05-23

[illegible]

; EARLIER APPLICATION NUMBER: 60/056,908  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/048,964  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/057,650  
; EARLIER FILING DATE: 1997-09-05  
; EARLIER APPLICATION NUMBER: 60/056,884  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/057,669  
; EARLIER FILING DATE: 1997-09-05  
; EARLIER APPLICATION NUMBER: 60/049,610  
; EARLIER FILING DATE: 1997-06-13  
; EARLIER APPLICATION NUMBER: 60/061,060  
; EARLIER FILING DATE: 1997-10-02

Query Match 1.9%; Score 7; DB 4; Length 61;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 TGTSSSQ 89  
|||||  
DB 40 TGTSSSQ 46

RESULT 11  
US-09-249-542-9  
; Sequence 9, Application US/09249542A  
; Patent No. 6392029

; GENERAL INFORMATION:  
; APPLICANT: Ludwig, Linda B.  
; TITLE OF INVENTION: HIV Chemokines  
; FILE REFERENCE: 11520.0122  
; CURRENT APPLICATION NUMBER: US/09/249,542A  
; CURRENT FILING DATE: 1999-02-12  
; EARLIER APPLICATION NUMBER: US 60/074,640  
; EARLIER FILING DATE: 1998-02-13  
; NUMBER OF SEQ ID NOS: 26

; SEQ ID NO 9  
; LENGTH: 81  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus

; FEATURE:  
; OTHER INFORMATION: HIV chemokine protein (2nd start site; (+1) ribosomal  
; OTHER INFORMATION: shift at nucleotide 201) from SF2 HIV  
US-09-249-542-9

Query Match 1.9%; Score 7; DB 4; Length 81;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 SGTSGTS 86  
|||||  
DB 58 SGTSGTS 64

RESULT 12  
US-09-249-542-8  
; Sequence 8, Application US/09249542A  
; Patent No. 6392029

; GENERAL INFORMATION:  
; APPLICANT: Ludwig, Linda B.  
; TITLE OF INVENTION: HIV Chemokines  
; FILE REFERENCE: 11520.0122  
; CURRENT APPLICATION NUMBER: US/09/249,542A  
; CURRENT FILING DATE: 1999-02-12  
; EARLIER APPLICATION NUMBER: US 60/074,640  
; EARLIER FILING DATE: 1998-02-13  
; NUMBER OF SEQ ID NOS: 26

; SEQ ID NO 8  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus

; FEATURE:

; OTHER INFORMATION: HIV chemokine protein (2nd start site; (-1) ribosomal  
; OTHER INFORMATION: shifts at nucleotides 200 and 265) from SF2 HIV  
US-09-249-542-8

Query Match 1.9%; Score 7; DB 4; Length 82;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 SGTSGTS 86  
|||||  
DB 59 SGTSGTS 65

RESULT 13  
US-09-036-113-2  
; Sequence 2, Application US/09036113  
; Patent No. 6077825  
; GENERAL INFORMATION:  
; APPLICANT: Cupp, Mary S.  
; APPLICANT: Cupp, Eddie W.  
; TITLE OF INVENTION: ANTITHROMBIN PROTEIN AND DNA SEQUENCES  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)  
; STREET: 3605 Glenwood Ave. Suite 310  
; CITY: Raleigh  
; STATE: NC  
; COUNTRY: USA  
; ZIP: 27622

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036.113  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spruill, W. Murray  
; REGISTRATION NUMBER: 32,943  
; REFERENCE/DOCKET NUMBER: 5721-5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919 420 2202  
; TELEFAX: 919 881 3175

; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 96 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-036-113-2

Query Match 1.9%; Score 7; DB 3; Length 96;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 CDRNEEL 355  
|||||  
DB 64 CDRNEEL 70

RESULT 14  
US-09-556-166-2  
; Sequence 2, Application US/09556166  
; Patent No. 6465214  
; GENERAL INFORMATION:  
; APPLICANT: Cupp, Mary S.  
; APPLICANT: Cupp, Eddie W.  
; TITLE OF INVENTION: ANTITHROMBIN PROTEIN AND DNA SEQUENCES  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)

```

; STREET: 3605 Glenwood Ave. Suite 310
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; ZIP: 27622
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/556,166
; FILING DATE: 21-APR-2000
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/036,113
; FILING DATE: 06-MAR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 5721-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919 420 2202
; TELEFAX: 919 881 3175
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-556-166-2

```

```

Query Match      1.9%; Score 7; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      349 CDRNEEL 355
        |||||||
DB      64 CDRNEEL 70

```

## RESULT 15

```

US-09-252-991A-29877
; Sequence 29877, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29877
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29877

```

```

Query Match      1.9%; Score 7; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      82 STGTSSS 88
        |||||||
DB      44 STGTSSS 50

```

## RESULT 16

```

US-09-249-542-16
; Sequence 16, Application US/09249542A
; Patent No. 6392029
; GENERAL INFORMATION:
; APPLICANT: Ludwig, Linda B.
; TITLE OF INVENTION: HIV Chemokines
; FILE REFERENCE: 11520.0122
; CURRENT APPLICATION NUMBER: US/09/249,542A
; CURRENT FILING DATE: 1999-02-12
; EARLIER APPLICATION NUMBER: US 60/074,640
; EARLIER FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 16
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; OTHER INFORMATION: HIV chemokine protein from YU2 strain with a (+1) and (-)
; OTHER INFORMATION: 1) ribosomal frameshift
US-09-249-542-16

```

```

Query Match      1.9%; Score 7; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      80 SGSTGTS 86
        |||||||
DB      89 SGSTGTS 95

```

## RESULT 17

```

US-09-249-542-17
; Sequence 17, Application US/09249542A
; Patent No. 6392029
; GENERAL INFORMATION:
; APPLICANT: Ludwig, Linda B.
; TITLE OF INVENTION: HIV Chemokines
; FILE REFERENCE: 11520.0122
; CURRENT APPLICATION NUMBER: US/09/249,542A
; CURRENT FILING DATE: 1999-02-12
; EARLIER APPLICATION NUMBER: US 60/074,640
; EARLIER FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 17
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; OTHER INFORMATION: HIV chemokine protein (1st start site; (-1) ribosomal
; OTHER INFORMATION: frameshift at nucleotide 133, (+1) ribosomal frameshift at
; OTHER INFORMATION: nucleotide 265) from YU2 HIV
US-09-249-542-17

```

```

Query Match      1.9%; Score 7; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      80 SGSTGTS 86
        |||||||
DB      89 SGSTGTS 95

```

## RESULT 18

```

US-09-249-542-18
; Sequence 18, Application US/09249542A
; Patent No. 6392029
; GENERAL INFORMATION:
; APPLICANT: Ludwig, Linda B.
; TITLE OF INVENTION: HIV Chemokines
; FILE REFERENCE: 11520.0122
; CURRENT APPLICATION NUMBER: US/09/249,542A
; CURRENT FILING DATE: 1999-02-12
; EARLIER APPLICATION NUMBER: US 60/074,640

```

```
; EARLIER FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 18
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: Xaa is any amino acid
; LOCATION: 28,37,47
; OTHER INFORMATION: HIV chemokine protein with read-through from YU2 HIV
US-09-249-542-18

Query Match      1.9%; Score 7; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      80 SGGTGTS 86
Db      89 SGGTGTS 95
      |||||

RESULT 19
US-09-249-542-7
; Sequence 7, Application US/09249542A
; Patent No. 6392029
; GENERAL INFORMATION:
; APPLICANT: Ludwig, Linda B.
; TITLE OF INVENTION: HIV Chemokines
; FILE REFERENCE: 11520.0122
; CURRENT APPLICATION NUMBER: US/09/249,542A
; CURRENT FILING DATE: 1999-02-12
; EARLIER APPLICATION NUMBER: US 60/074,640
; EARLIER FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 7
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; OTHER INFORMATION: HIV chemokine protein (1st start site; (+1) ribosomal
; OTHER INFORMATION: frameshift at nucleotide 133; and (-1) ribosomal shift at
; OTHER INFORMATION: nucleotide 265) from SF2 HIV
US-09-249-542-7

Query Match      1.9%; Score 7; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      80 SGGTGTS 86
Db      89 SGGTGTS 95
      |||||

RESULT 20
US-09-249-542-10
; Sequence 10, Application US/09249542A
; Patent No. 6392029
; GENERAL INFORMATION:
; APPLICANT: Ludwig, Linda B.
; TITLE OF INVENTION: HIV Chemokines
; FILE REFERENCE: 11520.0122
; CURRENT APPLICATION NUMBER: US/09/249,542A
; CURRENT FILING DATE: 1999-02-12
; EARLIER APPLICATION NUMBER: US 60/074,640
; EARLIER FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 10
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; OTHER INFORMATION: HIV chemokine protein (1st start site at nucleotide 52;
; OTHER INFORMATION: (+1) ribosomal frameshift at nucleotide 133, (-1) ribosomal
; OTHER INFORMATION: frameshift at nucleotide 265) from SF2 HIV
US-09-249-542-10

Query Match      1.9%; Score 7; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      80 SGGTGTS 86
Db      89 SGGTGTS 95
      |||||

RESULT 21
US-09-249-542-11
; Sequence 11, Application US/09249542A
; Patent No. 6392029
; GENERAL INFORMATION:
; APPLICANT: Ludwig, Linda B.
; TITLE OF INVENTION: HIV Chemokines
; FILE REFERENCE: 11520.0122
; CURRENT APPLICATION NUMBER: US/09/249,542A
; CURRENT FILING DATE: 1999-02-12
; EARLIER APPLICATION NUMBER: US 60/074,640
; EARLIER FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 11
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; OTHER INFORMATION: HIV chemokine protein (1st start site; (-1) ribosomal
; OTHER INFORMATION: frameshift at nucleotide 133, (+1) ribosomal frameshift at
; OTHER INFORMATION: nucleotide 265) from (Pt)CNS HIV
US-09-249-542-11

Query Match      1.9%; Score 7; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      80 SGGTGTS 86
Db      89 SGGTGTS 95
      |||||

RESULT 22
US-09-249-542-14
; Sequence 14, Application US/09249542A
; Patent No. 6392029
; GENERAL INFORMATION:
; APPLICANT: Ludwig, Linda B.
; TITLE OF INVENTION: HIV Chemokines
; FILE REFERENCE: 11520.0122
; CURRENT APPLICATION NUMBER: US/09/249,542A
; CURRENT FILING DATE: 1999-02-12
; EARLIER APPLICATION NUMBER: US 60/074,640
; EARLIER FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 14
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; OTHER INFORMATION: HIV chemokine protein (1st start site; (-1) ribosomal
; OTHER INFORMATION: frameshift at nucleotide 133, (+1) ribosomal frameshift at
; OTHER INFORMATION: nucleotide 265) from SF2 HIV
US-09-249-542-14

Query Match      1.9%; Score 7; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      80 SGGTGTS 86
Db      89 SGGTGTS 95
      |||||
```

Db 89 SGSTGTS 95

RESULT 23

US-09-249-542-15  
; Sequence 15, Application US/09249542A  
; Patent No. 6392029  
; GENERAL INFORMATION:  
; APPLICANT: Ludwig, Linda B.  
; TITLE OF INVENTION: HIV Chemokines  
; FILE REFERENCE: 11520.0122  
; CURRENT APPLICATION NUMBER: US/09/249,542A  
; CURRENT FILING DATE: 1999-02-12  
; EARLIER APPLICATION NUMBER: US 60/074,640  
; EARLIER FILING DATE: 1998-02-13  
; NUMBER OF SEQ ID NOS: 26  
; SEQ ID NO 15  
; LENGTH: 112  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus  
; FEATURE:  
; NAME/KEY: Xaa is any amino acid  
; LOCATION: 28,37,47  
; OTHER INFORMATION: HIV chemokine protein with read-through from SF2 HIV  
US-09-249-542-15

Query Match 1.9%; Score 7; DB 4; Length 112;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 SGSTGTS 86

Db 89 SGSTGTS 95

RESULT 24

US-09-249-542-19  
; Sequence 19, Application US/09249542A  
; Patent No. 6392029  
; GENERAL INFORMATION:  
; APPLICANT: Ludwig, Linda B.  
; TITLE OF INVENTION: HIV Chemokines  
; FILE REFERENCE: 11520.0122  
; CURRENT APPLICATION NUMBER: US/09/249,542A  
; CURRENT FILING DATE: 1999-02-12  
; EARLIER APPLICATION NUMBER: US 60/074,640  
; EARLIER FILING DATE: 1998-02-13  
; NUMBER OF SEQ ID NOS: 26  
; SEQ ID NO 19  
; LENGTH: 112  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus  
; FEATURE:  
; NAME/KEY: 28,37,97  
; LOCATION: Xaa is any amino acid  
; OTHER INFORMATION: HIV chemokine protein with read-through from ELI strain  
; OTHER INFORMATION: of HIV  
US-09-249-542-19

Query Match 1.9%; Score 7; DB 4; Length 112;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 GSTGTSS 87

Db 90 GSTGTSS 96

RESULT 25

US-09-249-542-24  
; Sequence 24, Application US/09249542A  
; Patent No. 6392029  
; GENERAL INFORMATION:

; APPLICANT: Ludwig, Linda B.  
; TITLE OF INVENTION: HIV Chemokines  
; FILE REFERENCE: 11520.0122  
; CURRENT APPLICATION NUMBER: US/09/249,542A  
; CURRENT FILING DATE: 1999-02-12  
; EARLIER APPLICATION NUMBER: US 60/074,640  
; EARLIER FILING DATE: 1998-02-13  
; NUMBER OF SEQ ID NOS: 26  
; SEQ ID NO 24  
; LENGTH: 112  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus  
; FEATURE:  
; NAME/KEY: Xaa is any amino acid  
; LOCATION: 97  
; OTHER INFORMATION: HIV chemokine protein with (-1) frameshift at nucleotide  
; OTHER INFORMATION: 133, (+1) frameshift at nucleotide 205, readthrough at  
; OTHER INFORMATION: nucleotides 341-343 from pHIVCAT  
US-09-249-542-24

Query Match 1.9%; Score 7; DB 4; Length 112;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 SGSTGTS 86

Db 89 SGSTGTS 95

RESULT 26

US-09-996-243-420  
; Sequence 420, Application US/09996243  
; Patent No. 6478825  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2730P1C13  
; CURRENT APPLICATION NUMBER: US/09/996,243  
; CURRENT FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24

;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089598
;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089599
;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089600
;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089653
;	PRIOR FILING DATE:	1998-06-17
;	PRIOR APPLICATION NUMBER:	60/089801
;	PRIOR FILING DATE:	1998-06-18
;	PRIOR APPLICATION NUMBER:	60/089907
;	PRIOR FILING DATE:	1998-06-18
;	PRIOR APPLICATION NUMBER:	60/089908
;	PRIOR FILING DATE:	1998-06-18
;	PRIOR APPLICATION NUMBER:	60/089947
;	PRIOR FILING DATE:	1998-06-19
;	PRIOR APPLICATION NUMBER:	60/089948
;	PRIOR FILING DATE:	1998-06-19
;	PRIOR APPLICATION NUMBER:	60/089952
;	PRIOR FILING DATE:	1998-06-19
;	PRIOR APPLICATION NUMBER:	60/090246
;	PRIOR FILING DATE:	1998-06-22
;	PRIOR APPLICATION NUMBER:	60/090252
;	PRIOR FILING DATE:	1998-06-22
;	PRIOR APPLICATION NUMBER:	60/090254
;	PRIOR FILING DATE:	1998-06-22
;	PRIOR APPLICATION NUMBER:	60/090349
;	PRIOR FILING DATE:	1998-06-23
;	PRIOR APPLICATION NUMBER:	60/090355
;	PRIOR FILING DATE:	1998-06-23
;	PRIOR APPLICATION NUMBER:	60/090429
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090431
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090435
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090444
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090445
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090472
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090535
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090540
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090542
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090557
;	PRIOR FILING DATE:	1998-06-24
;	PRIOR APPLICATION NUMBER:	60/090676
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090678
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090690
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090696
;	PRIOR FILING DATE:	1998-06-25
;	PRIOR APPLICATION NUMBER:	60/090862
;	PRIOR FILING DATE:	1998-06-26
;	PRIOR APPLICATION NUMBER:	60/090863
;	PRIOR FILING DATE:	1998-06-26
;	PRIOR APPLICATION NUMBER:	60/091360
;	PRIOR FILING DATE:	1998-07-01
;	PRIOR APPLICATION NUMBER:	60/091478
;	PRIOR FILING DATE:	1998-07-02
;	PRIOR APPLICATION NUMBER:	60/091544
;	PRIOR FILING DATE:	1998-07-0



; PRIOR APPLICATION NUMBER: 60/091519  
 ; PRIOR FILING DATE: 1998-07-02  
 ; PRIOR APPLICATION NUMBER: 60/091626  
 ; PRIOR FILING DATE: 1998-07-02  
 ; PRIOR APPLICATION NUMBER: 60/091633  
 ; PRIOR FILING DATE: 1998-07-02  
 ; PRIOR APPLICATION NUMBER: 60/091978  
 ; PRIOR FILING DATE: 1998-07-07  
 ; PRIOR APPLICATION NUMBER: 60/091982  
 ; PRIOR FILING DATE: 1998-07-07  
 ; PRIOR APPLICATION NUMBER: 60/092182  
 ; PRIOR FILING DATE: 1998-07-09

Query Match 1.9%; Score 7; DB 4; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGG 241  
 Db 83 GGGAGG 89

## RESULT 27

US-09-328-352-6723  
 ; Sequence 6723, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 6723  
 ; LENGTH: 160  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-6723

Query Match 1.9%; Score 7; DB 4; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 GAGGGPL 243  
 Db 128 GAGGGPL 134

## RESULT 28

US-08-403-852D-30  
 ; Sequence 30, Application US/08403852D  
 ; Patent No. 5891695  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blanc, Veronique  
 ; APPLICANT: Blanche, Francis  
 ; APPLICANT: Crouzet, Joel  
 ; APPLICANT: Jacques, Nathalie  
 ; APPLICANT: Lacroix, Patricia  
 ; APPLICANT: Thibaut, Denis  
 ; APPLICANT: Zagorec, Monique  
 ; APPLICANT: Debussche, Laurent  
 ; APPLICANT: De Crecy-Lagard, Valerie  
 ; TITLE OF INVENTION: Polypeptides Involved In The  
 ; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
 ; TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
 ; NUMBER OF SEQUENCES: 43  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
 ; STREET: 1300 I Street, N.W., Suite 700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA

; ZIP: 20005-3315  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/403.852D  
 ; FILING DATE: 10-MAY-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/FR 93/00923  
 ; FILING DATE: 25-SEP-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: FR 92/11441  
 ; FILING DATE: 25-SEP-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meyers, Kenneth J.  
 ; REGISTRATION NUMBER: 25,146  
 ; REFERENCE/DOCKET NUMBER: 03806.0054-00000  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 408-4000  
 ; TELEFAX: (202) 408-4400  
 ; INFORMATION FOR SEQ ID NO: 30:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 161 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 US-08-403-852D-30

Query Match 1.9%; Score 7; DB 2; Length 161;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 TSGSTGT 85  
 Db 81 TSGSTGT 87

## RESULT 29

US-08-510-646B-31  
 ; Sequence 31, Application US/08510646B  
 ; Patent No. 6077699  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blanc, Veronique  
 ; APPLICANT: Blanche, Francis  
 ; APPLICANT: Crouzet, Joel  
 ; APPLICANT: Jacques, Nathalie  
 ; APPLICANT: Lacroix, Patricia  
 ; APPLICANT: Thibaut, Denis  
 ; APPLICANT: Zagorec, Monique  
 ; APPLICANT: Debussche, Laurent  
 ; APPLICANT: De Crecy-Lagard, Valerie  
 ; TITLE OF INVENTION: Polypeptides Involved In The  
 ; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
 ; TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
 ; NUMBER OF SEQUENCES: 45  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
 ; STREET: 1300 I Street, N.W., Suite 700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3315  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/510.646B  
 ; FILING DATE: 03-AUG-1995  
 ; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/403,852  
;; FILING DATE: 10-MAY-1995  
;; PRIOR APPLICATION DATA: PCT/FR 93/00923  
;; APPLICATION NUMBER: PCT/FR 93/00923  
;; FILING DATE: 25-SEP-1993  
;; PRIOR APPLICATION DATA: FR 92/11441  
;; APPLICATION NUMBER: FR 92/11441  
;; FILING DATE: 25-SEP-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Meyers, Kenneth J.  
;; REGISTRATION NUMBER: 25,146  
;; REFERENCE/DOCKET NUMBER: 03806.0054-01000  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 408-4400  
;; TELEFAX: (202) 408-4400  
;; INFORMATION FOR SEQ ID NO: 31:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 161 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-08-510-646B-31

Query Match 1.9%; Score 7; DB 3; Length 161;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 TSGSTGT 85  
Db 81 TSGSTGT 87

RESULT 30  
US-09-231-818-30  
; Sequence 30, Application US/09231818  
; Patent No. 6171846  
; GENERAL INFORMATION:  
; APPLICANT: Blanc, Veronique  
; APPLICANT: Blanche, Francis  
; APPLICANT: Crouzet, Joel  
; APPLICANT: Jacques, Nathalie  
; APPLICANT: Lacroix, Patricia  
; APPLICANT: Thibaut, Denis  
; APPLICANT: Zagorec, Monique  
; APPLICANT: Debussche, Laurent  
; APPLICANT: De Crecy-Lagard, Valerie  
; TITLE OF INVENTION: Polypeptides Involved In The  
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/231,818  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/403,852  
; FILING DATE: 10-MAY-1995  
; APPLICATION NUMBER: PCT/FR 93/00923  
; FILING DATE: 25-SEP-1993  
; APPLICATION NUMBER: FR 92/11441

;; FILING DATE: 25-SEP-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Meyers, Kenneth J.  
;; REGISTRATION NUMBER: 25,146  
;; REFERENCE/DOCKET NUMBER: 03806.0054-00000  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 408-4000  
;; TELEFAX: (202) 408-4400  
;; INFORMATION FOR SEQ ID NO: 30:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 161 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-09-231-818-30

Query Match 1.9%; Score 7; DB 3; Length 161;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 TSGSTGT 85  
Db 81 TSGSTGT 87

RESULT 31  
US-09-562-737-112  
; Sequence 112, Application US/09562737  
; Patent No. 6428967  
; GENERAL INFORMATION:  
; APPLICANT: Herz, Joachim  
; APPLICANT: Gotthardt, Michael  
; TITLE OF INVENTION: LDL Receptor Signaling Pathways  
; FILE REFERENCE: UTSW0708  
; CURRENT APPLICATION NUMBER: US/09/562,737  
; CURRENT FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 112  
; LENGTH: 185  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: Sequence  
; US-09-562-737-112

Query Match 1.9%; Score 7; DB 4; Length 185;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLTVKTL 8  
Db 71 KLTVKTL 77

RESULT 32  
US-08-606-143-40  
; Sequence 40, Application US/08606143  
; Patent No. 5856155  
; GENERAL INFORMATION:  
; APPLICANT: Li, Min  
; TITLE OF INVENTION: COMPOUNDS AND RELATED METHODS FOR  
; TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH  
; TITLE OF INVENTION: COMPOUNDS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza, Suite 4900  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60601

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/606,143  
 ; FILING DATE: 23-FEB-1996  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kilyk Jr., John  
 ; REGISTRATION NUMBER: 30763  
 ; REFERENCE/DOCKET NUMBER: 71756  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (312) 616-5600  
 ; TELEFAX: (312) 616-5700  
 ; TELEX: 25-3533  
 ; INFORMATION FOR SEQ ID NO: 40:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 208 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-606-143-40

Query Match 1.9%; Score 7; DB 2; Length 208;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGAGG 240  
 |||||  
 Db 163 AGGAGG 169

## RESULT 33

; US-09-252-991A-25666  
 ; Sequence 25666, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 25666  
 ; LENGTH: 215  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-25666

Query Match 1.9%; Score 7; DB 4; Length 215;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGG 241  
 |||||  
 Db 136 GGGAGG 142

## RESULT 34

; US-08-506-553C-23  
 ; Sequence 23, Application US/08506553C  
 ; Patent No. 6120989  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Vornhagen, Rolf; Hinderer, Walter; Sonneborn, Han-H.;  
 ; APPLICANT: Plachter, Bodo; and Jahn, Gerhard

; TITLE OF INVENTION: ISOLATED HUMAN CYTOMEGALOVIRUS  
 ; TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 26  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Felfe & Lynch  
 ; STREET: 805 Third Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10022  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette  
 ; COMPUTER: IBM PS/2  
 ; OPERATING SYSTEM: PC-DOS  
 ; SOFTWARE: Wordperfect  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/506,553C  
 ; FILING DATE: 07/25/95  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mary Anne Schofield  
 ; REGISTRATION NUMBER: 36,669  
 ; REFERENCE/DOCKET NUMBER: LEDE 202-PFF/MAS  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 688-9200  
 ; TELEFAX: (212) 838-3884  
 ; INFORMATION FOR SEQ ID NO: 23:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 216 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Polypeptide  
 ; FRAGMENT TYPE: internal  
 ; US-08-506-553C-23

Query Match 1.9%; Score 7; DB 3; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGG 241  
 |||||  
 Db 163 GGGAGG 169

## RESULT 35

; US-09-107-532A-5362  
 ; Sequence 5362, Application US/09107532A  
 ; Patent No. 6583275  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 7310  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; CITY: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: PC  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,532A  
 ; FILING DATE: 30-Jun-1998  
 ; PRIOR APPLICATION DATA:

```
/ APPLICATION NUMBER: 60/085,598
/ FILING DATE: 14 May 1998
/ APPLICATION NUMBER: 60/051571
/ FILING DATE: July 2, 1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ariniello, Pamela Deneke
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-012
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-5277
/ INFORMATION FOR SEQ ID NO: 5362:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 226 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Enterococcus faecium
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATION 1...226
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5362:
US-09-107-532A-5362

Query Match 1.9%; Score 7; DB 4; Length 236;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 SNLSSGR 143
Db 53 SNLSSGR 59

RESULT 36
US-08-529-190B-1
/ Sequence 1, Application US/08529190B
/ Patent No. 5833991
/ GENERAL INFORMATION:
/ APPLICANT: Masucci, Maria G.
/ TITLE OF INVENTION: GLYCINE-CONTAINING SEQUENCES
/ TITLE OF INVENTION: CONFERRING INVISIBILITY TO THE IMMUNE SYSTEM
/ NUMBER OF SEQUENCES: 76
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Banner & Witcoff, Ltd.
/ STREET: One Financial Center
/ CITY: Boston
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02111
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: Wordperfect 6.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/529,190B
/ FILING DATE: 15-SEP-1995
/ CLASSIFICATION: 514
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: SE9501324-9
/ FILING DATE: 10-APR-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US08/522,595
/ FILING DATE: 01-SEP-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Williams, Ph.D., Kathleen A
/ REGISTRATION NUMBER: 34,380
/ REFERENCE/DOCKET NUMBER: 3255/53015
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617-345-9100
/ TELEFAX: 617-345-9111

/ APPLICATION NUMBER: 60/085,598
/ FILING DATE: 14 May 1998
/ APPLICATION NUMBER: 60/051571
/ FILING DATE: July 2, 1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ariniello, Pamela Deneke
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-012
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-5277
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 235 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: protein
US-08-529-190B-1

Query Match 1.9%; Score 7; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGG 240
Db 22 AGGAGG 28

RESULT 37
US-09-219-983A-12
/ Sequence 12, Application US/09219983A
/ Patent No. 6380159
/ GENERAL INFORMATION:
/ APPLICANT: Wolfner, Mariana
/ APPLICANT: Lung, Oliver
/ APPLICANT: Tram, Khanh-Uyen
/ TITLE OF INVENTION: GENES FOR MALE ACCESSORY GLAND PROTEINS IN DROSOPHILA
/ TITLE OF INVENTION: MELANOGASTER
/ FILE REFERENCE: 19603/1791
/ CURRENT APPLICATION NUMBER: US/09/219,983A
/ CURRENT FILING DATE: 1998-12-23
/ PRIOR APPLICATION NUMBER: 60/071,315
/ PRIOR FILING DATE: 1997-12-23
/ NUMBER OF SEQ ID NOS: 35
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 12
/ LENGTH: 238
/ TYPE: PRT
/ ORGANISM: Drosophila melanogaster
US-09-219-983A-12

Query Match 1.9%; Score 7; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GCGAGG 241
Db 22 GCGAGG 28

RESULT 38
US-08-414-926A-10
/ Sequence 10, Application US/08414926A
/ Patent No. 5721354
/ GENERAL INFORMATION:
/ APPLICANT: Spaete, Richard
/ APPLICANT: Cha, Tai-An
/ TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
/ NUMBER OF SEQUENCES: 27
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
/ STREET: 5 Palo Alto Square
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94306-2155
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/414,926A
```

; FILING DATE: March 31, 1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-494-7622  
; TELEFAX: 415-857-0663  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 240 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; CLONE: tol.04  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..240  
; OTHER INFORMATION: /label= UL136  
; US-08-414-926A-10

Query Match 1.9%; Score 7; DB 1; Length 240;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGG 240  
Db 186 AGGGAGG 192

## RESULT 39

US-08-926-922-10  
; Sequence 10, Application US/08926922  
; Patent No. 5925751  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Luann Cserr Attorney at Law  
; STREET: 750 Arimo Avenue  
; CITY: Oakland  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/926,922  
; FILING DATE: September 10, 1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR 11A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-834-1448  
; TELEFAX: 510-839-7810  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 240 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; CLONE: tol.04  
; FEATURE:

; NAME/KEY: Protein  
; LOCATION: 1..240  
; OTHER INFORMATION: /label= UL136  
; US-08-926-922-10

Query Match 1.9%; Score 7; DB 2; Length 240;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGG 240  
Db 186 AGGGAGG 192

## RESULT 40

US-09-253-682-10  
; Sequence 10, Application US/09253682  
; Patent No. 6040170  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; APPLICANT: Cha, Tai-An  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Luann Cserr Attorney at Law  
; STREET: 750 Arimo Avenue  
; CITY: Oakland  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/253,682  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/926,922  
; FILING DATE: September 10, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR 11A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-834-1448  
; TELEFAX: 510-839-7810  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 240 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; CLONE: tol.04  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..240  
; OTHER INFORMATION: /label= UL136  
; US-09-253-682-10

Query Match 1.9%; Score 7; DB 3; Length 240;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 234 AGGGAGG 240  
Db 186 AGGGAGG 192

## RESULT 41

US-09-527-657-10  
; Sequence 10, Application US/09527657  
; Patent No. 6291236  
; GENERAL INFORMATION:  
; APPLICANT: Spaete, Richard  
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Luann Cserr Attorney at Law  
; STREET: 750 Arimo Avenue  
; CITY: Oakland  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/527,657  
; FILING DATE: 17-Mar-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/926,922  
; FILING DATE: September 10, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: AVIR 11A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-834-1448  
; TELEFAX: 510-839-7810  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 240 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; CLONE: tol.04  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..240  
; OTHER INFORMATION: /label= ULL36  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-527-657-10  
  
Query Match 1.9%; Score 7; DB 3; Length 240;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 234 AGGAGG 240  
Db 186 AGGAGG 192  
|||||  
RESULT 42  
US-09-252-991A-22027  
; Sequence 22027, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27

US-09-527-657-10  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22027  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22027  
  
Query Match 1.9%; Score 7; DB 4; Length 242;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 235 GGGAGG 241  
Db 44 GGGAGG 50  
|||||  
RESULT 43  
US-09-266-965-123  
; Sequence 123, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456US1  
; CURRENT APPLICATION NUMBER: US/09/266,965  
; CURRENT FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 08/624,447  
; EARLIER FILING DATE: 1996-08-19  
; EARLIER APPLICATION NUMBER: PCT/US94/11279  
; EARLIER FILING DATE: 1994-10-06  
; EARLIER APPLICATION NUMBER: US 08/133,963  
; EARLIER FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 123  
; LENGTH: 254  
; TYPE: PRT  
; ORGANISM: Streptomyces lavendulae  
US-09-266-965-123  
  
Query Match 1.9%; Score 7; DB 4; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 201 GANTTDR 207  
Db 193 GANTTDR 199  
|||||  
RESULT 44  
US-09-219-983A-11  
; Sequence 11, Application US/09219983A  
; Patent No. 6380159  
; GENERAL INFORMATION:  
; APPLICANT: Wolfner, Mariana  
; APPLICANT: Lung, Oliver  
; APPLICANT: Tram, Khanh-Uyen  
; TITLE OF INVENTION: GENES FOR MALE ACCESSORY GLAND PROTEINS IN DROSOPHILA  
; FILE REFERENCE: 19603/1791  
; CURRENT APPLICATION NUMBER: US/09/219,983A  
; CURRENT FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/071,315  
; PRIOR FILING DATE: 1997-12-23  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 267  
; TYPE: PRT

```

; ORGANISM: Drosophila melanogaster
US-09-219-983A-11

Query Match      1.9%; Score 7; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      235 GGGAGGG 241
Db      50 GGGAGGG 56
|||||

RESULT 45
US-09-252-991A-16773
; Sequence 16773, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16773
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16773

Query Match      1.9%; Score 7; DB 4; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      234 AGGGAGG 240
Db      111 AGGGAGG 117
|||||

RESULT 46
US-08-506-553C-26
; Sequence 26, Application US/08506553C
; Patent No. 6120989
; GENERAL INFORMATION:
; APPLICANT: Vornhagen, Rolf; Hinderer, Walter; Sonneborn, Han-H.;
; APPLICANT: Plachter, Bodo; and Jahn, Gerhard
; TITLE OF INVENTION: ISOLATED HUMAN CYTOMEGALOVIRUS
; TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,553C
; FILING DATE: 07/25/95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LEADER 202-PFF/MAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 276 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Polypeptide
; FRAGMENT TYPE: internal
US-08-506-553C-26

Query Match      1.9%; Score 7; DB 3; Length 276;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      235 GGGAGGG 241
Db      163 GGGAGGG 169
|||||

RESULT 47
US-09-252-991A-19858
; Sequence 19858, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19858
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19858

Query Match      1.9%; Score 7; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      120 QPGLPLA 126
Db      262 QPGLPLA 268
|||||

RESULT 48
US-09-252-991A-26854
; Sequence 26854, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26854

```

; LENGTH: 307  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26854

Query Match 1.9%; Score 7; DB 4; Length 307;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 POPPVAP 115  
| | | | |  
DB 151 POPPVAP 157

## RESULT 49

US-09-252-991A-22411  
; Sequence 22411, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22411  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22411

Query Match 1.9%; Score 7; DB 4; Length 313;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
| | | | |  
DB 102 GGGAGGG 108

## RESULT 50

US-09-252-991A-22018  
; Sequence 22018, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22018  
; LENGTH: 322  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22018

Query Match 1.9%; Score 7; DB 4; Length 322;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 PFEKGEG 304  
| | | | |

DB 15 PFEKGEG 21

## RESULT 51

US-09-252-991A-33113  
; Sequence 33113, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 33113  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-33113

Query Match 1.9%; Score 7; DB 4; Length 323;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 HSNTPAT 96  
| | | | |  
DB 55 HSNTPAT 61

## RESULT 52

US-09-328-352-5177  
; Sequence 5177, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5177  
; LENGTH: 340  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5177

Query Match 1.9%; Score 7; DB 4; Length 340;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 EFLQLLN 296  
| | | | |  
DB 133 EFLQLLN 139

## RESULT 53

US-09-252-991A-17906  
; Sequence 17906, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788



; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17906  
; LENGTH: 352  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17906

Query Match 1.9%; Score 7; DB 4; Length 352;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AGGAGG 240  
Db 313 AGGAGG 319

RESULT 54  
US-09-252-991A-23504  
; Sequence 23504, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23504  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23504

Query Match 1.9%; Score 7; DB 4; Length 366;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGP 242  
Db 192 GGAGGGP 198

RESULT 55  
US-09-328-352-5630  
; Sequence 5630, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5630  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5630

Query Match 1.9%; Score 7; DB 4; Length 372;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 VQNDTI 23

Db 244 VQNDTI 250  
|||||

RESULT 56  
US-09-252-991A-24396  
; Sequence 24396, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24396  
; LENGTH: 441  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24396

Query Match 1.9%; Score 7; DB 4; Length 441;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GGAGGGP 242  
Db 188 GGAGGGP 194

RESULT 57  
US-08-821-355A-8  
; Sequence 8, Application US/08821355A  
; Patent No. 5851775  
; GENERAL INFORMATION:

; APPLICANT: Barker, Nick  
; APPLICANT: Clevers, Hans  
; APPLICANT: Korinek, Vladimir  
; APPLICANT: Morin, Patrice  
; APPLICANT: Kinzler, Kenneth  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Sparks, Andrew  
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC  
; TITLE OF INVENTION: Interact to Prevent Cancer  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/821,355A  
; FILING DATE: 20-MAR-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A  
; REGISTRATION NUMBER: 32,145  
; REFERENCE/DOCKET NUMBER: 1107.05064

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 97430 BMB UT  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5951775e  
US-08-821-355A-8

Query Match 1.9%; Score 7; DB 2; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
DB 8 GGGAGGG 14

RESULT 58  
US-09-003-687A-8  
Sequence 8, Application US/09003687A  
Patent No. 5998600  
GENERAL INFORMATION:

APPLICANT: Barker, Nick  
APPLICANT: Clevers, Hans  
APPLICANT: Korinek, Vladimir  
APPLICANT: Morin, Patrice  
APPLICANT: Kinzler, Kenneth  
APPLICANT: Vogelstein, Bert  
APPLICANT: Sparks, Andrew  
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC  
TITLE OF INVENTION: Interact to Prevent Cancer  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20001

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/003,687A  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/821,355  
FILING DATE: 20-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Kegan, Sarah A  
REGISTRATION NUMBER: 32,145  
REFERENCE/DOCKET NUMBER: 1107.05064  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 97430 BMB UT

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5998600e  
US-09-003-687A-8

Query Match 1.9%; Score 7; DB 2; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
DB 8 GGGAGGG 14

RESULT 59  
US-09-136-605-8  
Sequence 8, Application US/09136605A  
Patent No. 6140052  
GENERAL INFORMATION:  
APPLICANT: He, Tong-Chuan  
APPLICANT: Kinzler, Kenneth  
APPLICANT: Vogelstein, Bert  
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to  
TITLE OF INVENTION: Prevent Cancer  
FILE REFERENCE: 1107.75741  
CURRENT APPLICATION NUMBER: US/09/136,605A  
CURRENT FILING DATE: 1998-08-20  
EARLIER FILING DATE: 1997-03-20  
EARLIER APPLICATION NUMBER: 09/003,687  
EARLIER FILING DATE: 1998-01-06  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 8  
LENGTH: 486  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-136-605-8

Query Match 1.9%; Score 7; DB 3; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
DB 8 GGGAGGG 14

RESULT 60  
US-08-821-355A-9  
Sequence 9, Application US/08821355A  
Patent No. 5851775  
GENERAL INFORMATION:  
APPLICANT: Barker, Nick  
APPLICANT: Clevers, Hans  
APPLICANT: Korinek, Vladimir  
APPLICANT: Morin, Patrice  
APPLICANT: Kinzler, Kenneth  
APPLICANT: Vogelstein, Bert  
APPLICANT: Sparks, Andrew  
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC  
TITLE OF INVENTION: Interact to Prevent Cancer  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20001

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/821,355A  
FILING DATE: 20-MAR-1997

```
;
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32,145
; REFERENCE/DOCKET NUMBER: 1107.05064
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9299
; TELEFAX: 202-508-9299
; TELEX: 97430 BMB UT
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 511 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5851775e
; US-08-821-355A-9

Query Match 1.9%; Score 7; DB 2; Length 511;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241
Db 8 GGGAGGG 14

RESULT 61
US-09-003-687A-9
; Sequence 9, Application US/09003687A
; Patent No. 5998600
; GENERAL INFORMATION:
; APPLICANT: Barker, Nick
; APPLICANT: Clevers, Hans
; APPLICANT: Korinek, Vladimir
; APPLICANT: Morin, Patrice
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; APPLICANT: Sparks, Andrew
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
; TITLE OF INVENTION: Interact to Prevent Cancer
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-SEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/003,687A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: 08/821,355
; FILING DATE: 20-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32,145
; REFERENCE/DOCKET NUMBER: 1107.05064
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 97430 BMB UT
; INFORMATION FOR SEQ ID NO: 9:
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 511 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5998600e
; US-09-003-687A-9

Query Match 1.9%; Score 7; DB 2; Length 511;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241
Db 8 GGGAGGG 14

RESULT 62
US-09-136-605-9
; Sequence 9, Application US/09136605A
; Patent No. 6140052
; GENERAL INFORMATION:
; APPLICANT: He, Tong-Chuan
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to
; TITLE OF INVENTION: Prevent Cancer
; FILE REFERENCE: 1107.75741
; CURRENT APPLICATION NUMBER: US/09/136,605A
; CURRENT FILING DATE: 1998-08-20
; EARLIER APPLICATION NUMBER: 08/821,355
; EARLIER FILING DATE: 1997-03-20
; EARLIER APPLICATION NUMBER: 09/003,687
; EARLIER FILING DATE: 1998-01-06
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Fast-SEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-136-605-9

Query Match 1.9%; Score 7; DB 3; Length 511;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241
Db 8 GGGAGGG 14

RESULT 63
US-09-252-991A-18266
; Sequence 18266, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18266
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-18266
```

Query Match 1.9%; Score 7; DB 4; Length 521;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 GGGAGGG 241  
DB 152 GGGAGGG 158  
|||||

RESULT 64  
US-08-687-590-27  
; Sequence 27, Application US/08687590  
; Patent No. 6255070  
; GENERAL INFORMATION:  
; APPLICANT: Willison, Keith Robert  
; APPLICANT: Kubota, Hiroshi  
; APPLICANT: Ashworth, Alan  
; TITLE OF INVENTION: Folding Proteins  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/687,590  
; FILING DATE: 31-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB95/00192  
; FILING DATE: 31-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9401791.0  
; FILING DATE: 31-JAN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9418234.2  
; FILING DATE: 09-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 084619-0000000US  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 539 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-687-590-27

Query Match 1.9%; Score 7; DB 3; Length 539;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 MLVELSK 275  
DB 90 MLVELSK 96  
|||||

RESULT 65  
US-09-702-705-326  
; Sequence 326, Application US/09702705  
; Patent No. 6504010  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary  
; APPLICANT: Vedvick, Tom  
; APPLICANT: Carter, Darriick  
; APPLICANT: Retter, Marc  
; APPLICANT: Mannion, Jane  
; APPLICANT: Fan, Liqun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.478C14  
; CURRENT APPLICATION NUMBER: US/09/702,705  
; NUMBER OF SEQ ID NOS: 1833  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 326  
; LENGTH: 539  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-702-705-326

Query Match 1.9%; Score 7; DB 4; Length 539;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 MLVELSK 275  
DB 90 MLVELSK 96  
|||||

RESULT 66  
US-09-736-457-326  
; Sequence 326, Application US/09736457  
; Patent No. 6509448  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Lodes, Michael A.  
; APPLICANT: Fanger, Gary  
; APPLICANT: Vedvick, Tom  
; APPLICANT: Carter, Darriick  
; APPLICANT: Retter, Marc  
; APPLICANT: Mannion, Jane  
; APPLICANT: Fan, Liqun  
; APPLICANT: Wang, Aijun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.478C15  
; CURRENT APPLICATION NUMBER: US/09/736,457  
; NUMBER OF SEQ ID NOS: 1864  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 326  
; LENGTH: 539  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-736-457-326

Query Match 1.9%; Score 7; DB 4; Length 539;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 MLVELSK 275  
DB 90 MLVELSK 96  
|||||

RESULT 67  
US-09-067-351-1  
; Sequence 1, Application US/09067351  
; Patent No. 5994081  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.

APPLICANT: Baughn, Mariah  
TITLE OF INVENTION: HUMAN KERATINS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
FILING DATE: Herewith  
APPLICATION NUMBER: US/09/067,351  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PP-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 546 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PANCUT02  
CLONE: 1467090  
US-09-067-351-1

Query Match 1.9% Score 7; DB 2; Length 546;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241  
Db 108 GGGAGGG 114

RESULT 68  
US-09-360-490-1  
Sequence 1, Application US/09360490  
Patent No. 6221843  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Baughn, Mariah  
TITLE OF INVENTION: HUMAN KERATINS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
FILING DATE:  
APPLICATION NUMBER: US/09/360,490  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/067,351  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PP-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 546 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PANCUT02  
CLONE: 1467090  
US-09-360-490-1

Query Match 1.9% Score 7; DB 3; Length 546;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 GGGAGGG 241  
Db 108 GGGAGGG 114

RESULT 69  
US-09-413-814-89  
Sequence 89, Application US/09413814  
Patent No. 6225064  
GENERAL INFORMATION:  
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
APPLICANT: Bristol-Myers Squibb, Co.  
APPLICANT: Bayer, Stefan  
APPLICANT: Bloecker, Helmut  
APPLICANT: Brandt, Petra  
APPLICANT: Cino, Paul M  
APPLICANT: Dougherty, Brian A  
APPLICANT: Goldberg, Steven L  
APPLICANT: Hofle, Gerhard  
APPLICANT: Mueller, Joachim  
APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
FILE OF INVENTION: heteropolyketide compounds  
FILE REFERENCE: PCT/US 99/23535  
CURRENT APPLICATION NUMBER: US/09/413,814  
CURRENT FILING DATE: 1999-10-07  
EARLIER APPLICATION NUMBER: DE 198 46 493.2  
EARLIER FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 89  
LENGTH: 590  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-413-814-89

Query Match 1.9% Score 7; DB 3; Length 590;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 79 TSGSTGT 85  
Db 241 TSGSTGT 247

RESULT 70  
US-09-413-814-76  
Sequence 76, Application US/09413814

```
/ Patent No. 6225064
/ GENERAL INFORMATION:
/ APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
/ APPLICANT: Bristol-Myers Squibb, Co.
/ APPLICANT: Beyer, Stefan
/ APPLICANT: Bloeker, Helmut
/ APPLICANT: Brandt, Petra
/ APPLICANT: Cino, Paul M
/ APPLICANT: Dougherty, Brian A
/ APPLICANT: Goldberg, Steven L
/ APPLICANT: Hofle, Gerhard
/ APPLICANT: Mueller, Joachim
/ APPLICANT: Reichenbach, Hans
/ TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
/ FILE OF INVENTION: heteropolyketide compounds
/ FILE REFERENCE: PCT/US 99/23535
/ CURRENT APPLICATION NUMBER: US/09/413,814
/ CURRENT FILING DATE: 1999-10-07
/ EARLIER APPLICATION NUMBER: DE 198 46 493.2
/ EARLIER FILING DATE: 1998-10-09
/ NUMBER OF SEQ ID NOS: 107
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 76
/ LENGTH: 591
/ TYPE: PRT
/ ORGANISM: Sorangium cellulosum
US-09-413-814-76

Query Match          1.9%; Score 7; DB 3; Length 591;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 TSGSTGT 85
   |||||
Db 242 TSGSTGT 248

RESULT 71
US-09-252-991A-20134
/ Sequence 20134, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 20134
/ LENGTH: 609
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20134

Query Match          1.9%; Score 7; DB 4; Length 609;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 LIENHD 289
   |||||
Db 157 LIENHD 163

RESULT 72
US-08-894-454-163
/ Sequence 163, Application US/08894454
/ Patent No. 6544784
/ GENERAL INFORMATION:
/ APPLICANT: VAN DEN VEN, W.J.M.
/ APPLICANT: SCHOENMAKERS, H.F.P.M.
/ TITLE OF INVENTION: MULTIPLE-TUMOR ABERRANT GROWTH
/ FILE OF INVENTION: GENES
/ NUMBER OF SEQUENCES: 164
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: The Webb Law Firm
/ STREET: 700 Koppers Building, 436 Seventh Avenue
/ CITY: Pittsburgh
/ STATE: PA
/ COUNTRY: USA
/ ZIP: 15219-1818
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/894,454
/ FILING DATE: 15-AUG-1997
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/EP/00716
/ FILING DATE: 19-FEB-1996
/ APPLICATION NUMBER: 95200390.3
/ FILING DATE: 17-FEB-1995
/ APPLICATION NUMBER: 95201951.1
/ FILING DATE: 14-JUL-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Johnson, Barbara E
/ REGISTRATION NUMBER: 31,198
/ REFERENCE/DOCKET NUMBER: 702-971100
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 412-471-8815
/ TELEFAX: 412-471-4094
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 163:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 612 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
US-08-894-454-163

Query Match          1.9%; Score 7; DB 4; Length 612;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 GGEGDFL 307
   |||||
Db 62 GGEGDFL 68

RESULT 73
US-09-328-352-4474
/ Sequence 4474, Application US/09328352
/ Patent No. 6562958
/ GENERAL INFORMATION:
/ APPLICANT: Gary L. Breton et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
/ FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC99-03PA
/ CURRENT APPLICATION NUMBER: US/09/328,352
/ CURRENT FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 8252
/ SEQ ID NO 4474
/ LENGTH: 615
/ TYPE: PRT
/ ORGANISM: Acinetobacter baumannii
US-09-328-352-4474

Query Match          1.9%; Score 7; DB 4; Length 615;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 79 TSGSTGT 85  
| | | | |  
Db 164 TSGSTGT 170

RESULT 74  
US-09-345-473E-46  
; Sequence 46, Application US/09345473E  
; Patent No. 6558903  
; GENERAL INFORMATION:  
; APPLICANT: Hodge, Martin  
; TITLE OF INVENTION: NO. 6558903el Kinases and Uses Thereof  
; FILE REFERENCE: 35800/183781  
; CURRENT APPLICATION NUMBER: US/09/345,473E  
; CURRENT FILING DATE: 1999-06-30  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 46  
; LENGTH: 627  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-345-473E-46

Query Match 1.9%; Score 7; DB 4; Length 627;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 GGAGGGP 242  
| | | | |  
Db 22 GGAGGGP 28

RESULT 75  
US-09-345-473E-47  
; Sequence 47, Application US/09345473E  
; Patent No. 6558903  
; GENERAL INFORMATION:  
; APPLICANT: Hodge, Martin  
; TITLE OF INVENTION: NO. 6558903el Kinases and Uses Thereof  
; FILE REFERENCE: 35800/183781  
; CURRENT APPLICATION NUMBER: US/09/345,473E  
; CURRENT FILING DATE: 1999-06-30  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 47  
; LENGTH: 627  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-345-473E-47

Query Match 1.9%; Score 7; DB 4; Length 627;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 GGAGGGP 242  
| | | | |  
Db 22 GGAGGGP 28

Search completed: December 17, 2003, 06:36:30  
Job time : 14.8538 secs

09/805-550

2. US-09-805-550-2 (1-405)  
US-09-805-550-4 Sequence 4, Application US/09805550

Sequence 4, Application US/09805550

GENERAL INFORMATION:

APPLICANT: Mahajan, Pramod B.  
APPLICANT: Tagliani, Laura  
TITLE OF INVENTION: Rad23 Genes and Uses Thereof  
FILE REFERENCE: 0964D  
CURRENT APPLICATION NUMBER: US/09/805,550  
CURRENT FILING DATE: 2001-03-13  
PRIOR APPLICATION NUMBER: 09/413,574  
PRIOR FILING DATE: 1999-10-06  
PRIOR APPLICATION NUMBER: 60/109,728  
PRIOR FILING DATE: 1998-11-23

NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4

LENGTH: 368

TYPE: PRT

ORGANISM: Zea mays

Initial Score = 76 Optimized Score = 125 Significance = -0.70  
Residue Identity = 34% Matches = 137 Mismatches = 215  
Gaps = 40 Conservative Substitutions = 0

```

10      20      30      40      50      60      70
M K L N V K T L G T N F E I E A S P D A S V A D V K R I I E T T Q G S Y R A D Q Q M L I Y G G K I L K D E T T L E S N G V A E N S F L V I
      |      |      |      |      |      |      |
      X      10      20      30      40
M K L T V K T L G T H F E I R V Q P N D T I M A V K N I E --- E I Q G K D S Y P W G Q Q
      X

80      90      100     110     120     130     140
M L S K A K A S S G A S T A T T A K - - A P A T L A Q P A A P A A S V A R T P T Q A P V A T A - E T A P P S V Q P Q A A P A A T V A A T
      |      |      |      |      |      |      |
      |      |      |      |      |      |      |
L L I F N G K V L K D E S T L E E N K V N E D G F L V M L S K G T S G S T G T S S Q H S N T P A T R Q A P P L E A P Q Q A P P P V A - -
      |      |      |      |      |      |      |
      |      |      |      |      |      |      |
50      60      70      80      90     100     110
D D A D V Y S Q A S N L V F G N N L E Q T I Q I L D M G G T W E R D T V V R A L - - - - - R A A Y N N P E R A I D Y L I S G I P E N V E
      |      |      |      |      |      |      |
      |      |      |      |      |      |      |
-- P I T T S Q P E G L P A Q A P N T H D N A A N L L S G - - - R N V D T I I N Q L M E M G G S W D K V Q R A L R A A Y N N P E R A V E
      |      |      |      |      |      |      |
      |      |      |      |      |      |      |
120     130     140     150     160     170     180
----- A Q P V - - - A R A P A G Q T T N Q Q A S P A Q P A V A L F V Q P S P A S A G P N A N P L N L F P Q G V P S G G S N P G V V F G A G
      |      |      |      |      |      |      |
      |      |      |      |      |      |      |
Y L Y S G I P V T A E I A V R I G G Q G A N T T D R A P T G E A - - - - - G L S G I P N T A P L D L F P Q - - - - - G A S N A G - - G G A G
      |      |      |      |      |      |      |
      |      |      |      |      |      |      |
190     200     210     220     230
S G A L D A L R O L P O F A L L Q L V A N P Q I L Q M L Q E L G K Q N P Q I L R L I O E N Q A B F L R L V N S P E G G P G N I L G Q L
      |      |      |      |      |      |      |
      |      |      |      |      |      |      |
G G P L D F L R N N P O F A V R E N V H T N P Q I L Q P M L V E L S K Q N P Q I L R L I E E N H D E F L Q L L N E P F E G G E G D F L D Q P E
      |      |      |      |      |      |      |
      |      |      |      |      |      |      |
240     250     260     270     280     290     300     310
A A A V P Q T L T V T P E E R A I Q R L E G M G F N R E L V I E V F A C N K D E E L T A N Y L L D H G H F D Q Q Q
      |      |      |      |      |      |      |
      |      |      |      |      |      |      |
E D E M P H A I S V T P E E Q E A I G R L E S M G F D R A R V I E A F L A C D R N E E L A A N Y L L E H A G E D
      |      |      |      |      |      |      |
      |      |      |      |      |      |      |
320     330     340     350     360     370     380     390     400
X
```



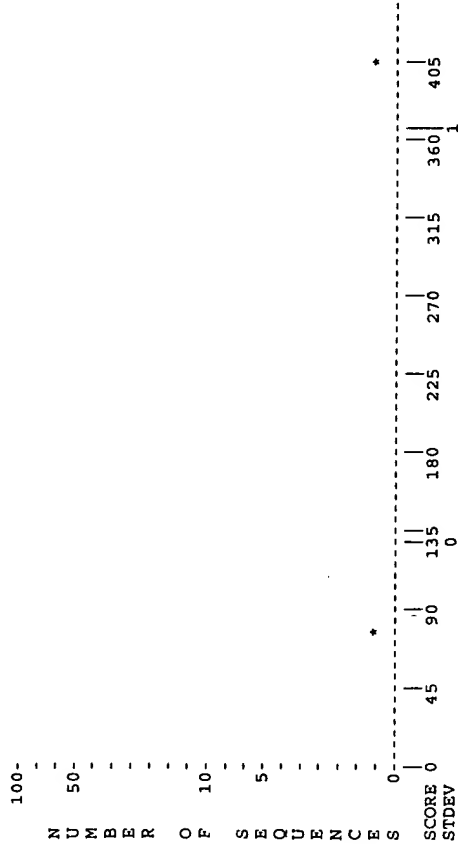
> O <  
O| O IntelliGenetics  
> O <

FastDB - Fast Pairwise Comparison of Sequences  
Release 5.4

Results file us-09-805-550-2.res made by jdelaval on Wed 17 Dec 103 6:22:09-PST.

Query sequence being compared: US-09-805-550-2 (1-405)  
Number of sequences searched: 2  
Number of scores above cutoff: 2

Results of the initial comparison of US-09-805-550-2 (1-405) with:  
File : US09805550.pep



PARAMETERS

Similarity matrix Unitary  
Mismatch penalty 1  
Gap penalty 1.00  
Gap size penalty 0.05  
Cutoff score 0  
Randomization group 0

SEARCH STATISTICS

Scores: Mean 240 Median 77 Standard Deviation 232.64  
Times: CPU 00:00:00.00 Total Elapsed 00:00:00.00

Number of residues: 773  
Number of sequences searched: 2  
Number of scores above cutoff: 2

The scores below are sorted by initial score.  
Significance is calculated based on initial score.  
A 100% identical sequence to the query sequence was found:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame

1. US-09-805-550-2 Sequence 2, Application US 405 405 405 0.71 0
- The list of other best scores is:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
2. US-09-805-550-4 Sequence 4, Application US 368 76 125 -0.70 0					

1. US-09-805-550-2 (1-405)  
US-09-805-550-2 Sequence 2, Application US/09805550

Sequence 2, Application US/09805550  
GENERAL INFORMATION:  
APPLICANT: Mahajan, Pramod B.  
TITLE OF INVENTION: Rad23 Genes and Uses Thereof  
FILE REFERENCE: 0964D  
CURRENT APPLICATION NUMBER: US/09/805,550  
CURRENT FILING DATE: 2001-03-13  
PRIOR APPLICATION NUMBER: 09/413,574  
PRIOR FILING DATE: 1999-10-06  
PRIOR APPLICATION NUMBER: 60/109,728  
PRIOR FILING DATE: 1998-11-23  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 405  
TYPE: PRT  
ORGANISM: Zea mays

Initial Score = 405 Optimized Score = 405 Significance = 0.71  
Residue Identity = 100% Matches = 405 Mismatches = 0  
Gaps = 0 Conservative Substitutions = 0

X 10 20 30 40 50 60 70  
MKNVTKLGTNFEIEASPDASVADVKKRIETTCQSTYRADQQMLIYQGIKDKDETTLESNGVAENSLVI  
MKNVTKLGTNFEIEASPDASVADVKKRIETTCQSTYRADQQMLIYQGIKDKDETTLESNGVAENSLVI  
X 10 20 30 40 50 60 70  
80 90 100 110 120 130 140  
MLSKAKASSSGASTATTAKAPATLAQPAAPVAPASVARTPTQAPVATAETAPPVQPOQAAATVAATDDA  
MLSKAKASSSGASTATTAKAPATLAQPAAPVAPASVARTPTQAPVATAETAPPVQPOQAAATVAATDDA  
80 90 100 110 120 130 140  
150 160 170 180 190 200 210  
DVYQAASNLVFGNNLEQTIQIILDMGGTWERDVTVRALRAAYNNPERAIDYLSGIPENVEAQPVARAPA  
DVYQAASNLVFGNNLEQTIQIILDMGGTWERDVTVRALRAAYNNPERAIDYLSGIPENVEAQPVARAPA  
150 160 170 180 190 200 210  
220 230 240 250 260 270 280  
AGQOTNQQAASPAQPAVALPVQSPASAGPNANPLNLFPGVPSGSGNPGVVPAGSGALDALROLPOQAL  
AGQOTNQQAASPAQPAVALPVQSPASAGPNANPLNLFPGVPSGSGNPGVVPAGSGALDALROLPOQAL  
220 230 240 250 260 270 280  
290 300 310 320 330 340 350 360  
LQLVQANFQIQLQPMQLQELGKONPQILRIQENQABFLRLVNESPGGPGNLTGALAAVPTLTTPERE  
LQLVQANFQIQLQPMQLQELGKONPQILRIQENQABFLRLVNESPGGPGNLTGALAAVPTLTTPERE  
290 300 310 320 330 340 350 360  
370 380 390 400 X  
AIIQLEGMGFNRELVLVEFFACNCKDEELTANYLLDHGHEFDQOQ  
AIIQLEGMGFNRELVLVEFFACNCKDEELTANYLLDHGHEFDQOQ  
370 380 390 400 X

> O <  
O | O IntelliGenetics  
> O <

# GENALIGN - Multiple Sequence Alignment Program Release 5.4

Wed 17 Dec 103 6:20:37-PST

## Solution Parameters:

Amino Alphabet = Identity  
Output line length = 80  
Compress = Off  
Histogram = Off  
Randomization = Off  
  
AMINO-Res-length = 2  
Deletion-weight = 5.00  
Length-factor = 0  
Matching-weight = 1.00  
NUCLEIC-Res-length = 4  
Spread-factor = 50

## Clustered order of selected sequences:

1. US-09-805-550-2 (1-405)
2. US-09-805-550-4 (1-368)

## Region Alignment: (listed in Clustered order)

```

US-09-805- 1 MKLnVKTlKGtNFIEasPdasvadVKrIEtTQqstYradQQmLIyqGKiLKDEtTLES
US-09-805- 1 MKLcVKTlKGtNFIEirvPndtImaVKNIeIqGkdsYpwgQQLlInGKvLKDEsTLBe
consensus  MKL-VKTlKGT-FEI---P-----VK--IE--QG---Y---QQ-LI--GK-LKDE-TLE-

US-09-805- 62 NgVaEnsfLVImlSKakassGstattakapatlaqPAapvAPaasvartPtQAPvatae
US-09-805- 62 NkVnEdgFLVvWMLSK gktSG ST gssssghentPAtzqAP pleapQAP qppv
consensus  N-V-E--FLV-MLSKak---SGaStat-----PA-----APaas-----P-QAPV----

US-09-805- 123 tappsvQPqaPaatvaatddadvysqAASNLvfgmNleqtIqildMGGGtWerDtVvRA
US-09-805- 114 apittsQegilPA qpnthdnAASNLlGrNvdtIInQlmeWGGGsWdkDkVqRA
consensus  -----QP---Paatvaat-----AASNL--G-N-----I-Q---MGGG-W--D-V-RA

US-09-805- 184 LRAAYNNPERAidYlSGIPenveaqPvaraPaasGQqtnqqaasPqapavalPvqpspsa
US-09-805- 169 LRAAYNNPERAveYlSGIP vtaeiavPigGQ gAnttdraPtgeaglsG
consensus  LRAAYNNPERA--YlSGIPenve-----P--GQqtnqqaas-A-----P-----S-

US-09-805- 245 gPnanPLaLPQgPpsgSNpvyvGAGsGaLdAlRglPQFQAIlqlvqanPQILQPMlQe
US-09-805- 218 iPNtaPLdLPQg aSN aggGAGgPlDfLRnnPQFQAvremVhtNPQILQPMlVe
consensus  -PN--PL-LFPQGVpssg-SNPg---GAG-G-LD-LR--PQFQA----V--NPQILQPMl-E

US-09-805- 306 lGKNPQILRLlIqNgaeFLvNNEspEGGpGgnilglaaavPqtltVTPEE-EAIGRLS
US-09-805- 273 LsKNPQILRLlIeNhdeFLdlNNEpEGGgGdldqpeedemPhaisVTPEE-EAIGRLS
consensus  L-KQNpQILRLlI-EN--EFL-L-NE--EGG-G-----P-----VTPEE-EAIGRLS

```

```

US-09-805- 367 gMGfNrElVlEvFfACnkdbEELtANYLLdHghsfDdqqg
US-09-805- 334 eMGfDrarViEaFlACdrnEELaANYLLLeHagReD
consensus  -MGf-R--V-E-P-AC---EEL-ANYLL-H--E-Ddqqg

```

Alignment score = 7.00

## Scoring matrix:

	1	2
1		-1
2		

-----